Pharmakophorbasierte Leitstruktursuche in chemischen Fragmenträumen mit stochastischen Methoden

Dissertation
zur Erlangung des akademischen Grades
Dr. rer. nat.
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Universität Hamburg

eingereicht beim Fachbereich Informatik von

Tobias Lippert
aus Darmstadt

Hamburg, Dezember 2011

Erster Gutachter:	Prof. Dr. Matthias Rarey, Universität Hamburg
Zweiter Gutachter:	Prof. Dr. Andrew Torda, Universität Hamburg
Tag der Disputation:	24.08.2012

Abstract

Virtual *de novo* design is employed in pharmaceutical research to help in the pursuit of new ligands for a chosen target by creating (partially) novel molecular scaffolds. The methods which are applied in this field recently shifted from atom-based to fragment-based approaches, where so-called chemical fragment spaces are used to generate molecular structures. In this work, a new method for *de novo* design is presented under the name „Qsearch". It makes use of stochastic methods to navigate chemical fragment spaces under the constraints of three-dimensional pharmacophores, and augments the fragment space approach by modifying the input such that the fragment space is adapted intelligently for a specific pharmacophore.

Pharmacophores are an established method in medicinal chemistry which allow to characterize key interactions, which are essential for molecular recognition of small molecules by their biological target in an intuitive way. A particular strength of them is that they represent chemical information independently from specific molecular motifs. This is especially useful for *de novo* design methods, since one of their key goals is to incorporate as much novelty into the generated molecules as possible.

The input for Qsearch consists of a pharmacophore and a fragment space. The fragment space is navigated with a Simulated Annealing procedure, which iteratively creates molecules by changing their constituting fragments. Every created molecule is subjected to a three-dimensional alignment to the input pharmacophore. In a following undecoration step, the complexity of the generated molecules is reduced by clipping terminal functional groups and removing hetero atoms in aromatic heterocycles—provided that the pharmacophore allows for these changes. This step allows the algorithm to access molecules which do not result directly from connecting the input fragments, and to extend the number of accessible molecules in a meaningful way. Furthermore, slight modifications to the workflow of Qsearch enable it to grow complete molecules from anchor fragments, which is a common task in medicinal chemistry.

Qsearch has been tested in several scenarios and has been shown to be able to generate relevant ideas for novel molecular scaffolds.

Zusammenfassung

de novo Design wird in der Wirkstoffforschung dazu eingesetzt, um neuartige molekulare Grundgerüste für eine bestimmte Zielstruktur zu erzeugen. Dabei hat sich die Methodik zur Generierung der Grundgerüste in den letzten Jahre von einem atombasierten zu einem fragmentbasierten Ansatz verschoben. In dieser Arbeit wird unter dem Namen „Qsearch“ eine neu entwickelte *de novo* Methode vorgestellt, die den fragmentbasierten Ansatz erweitert und die Eingabe intelligent so modifiziert, dass der zugängliche chemische Raum für die jeweilige Fragestellung dynamisch angepasst wird. Dabei wird die Suche in den Fragmenträumen mithilfe von stochastischen Methoden vorangetrieben.

Die Suchanfragen, die in Qsearch verwendet werden, um die Anforderungen an die zu generierenden Grundgerüste zu formulieren, sind dreidimensionale Pharmakophore. Pharmakophore sind eine etablierte Methode in der medizinischen Chemie, die es erlauben, die Schlüsselinteraktionen, die für die *molekulare Erkennung* von kleinen Wirkstoffmolekülen durch biologische Makromoleküle essentiell sind, auf eine intuitive Art und Weise zu charakterisieren. Eine ihrer besonderen Stärken ist, dass sie die relevanten chemischen Eigenschaften unabhängig von konkreten molekularen Strukturmotiven darstellen, was für die Verwendung in *de novo* Methoden vorteilhaft ist, da dort möglichst neuartige Moleküle als Ergebnis gewünscht sind.

Die Eingabe für Qsearch besteht aus einem Pharmakophor und einem Fragmentraum. Der Fragmentraum wird mit einer Simulated Annealing Optimierungsprozedur durchsucht, wobei iterativ Moleküle durch Veränderung auf Fragmentebene generiert und danach dreidimensional an dem Pharmakophor ausgerichtet werden. In einem darauf folgenden Entdekorierungsschritt wird die Komplexität der Moleküle dadurch reduziert, dass endständige funktionelle Gruppen eliminiert werden, und die Anzahl von Heteroatomen in Heteroaromaten verringert wird, wenn dies der Erfüllung des Pharmakophors nicht entgegensteht. Dieser Schritt erlaubt es, Zugang zu einem erweiterten Suchraum zu erlangen, der für die Erfüllung des Eingabepharmakophors relevant ist. Zusätzlich besteht die Möglichkeit, mit einer leichten Abwandlung der Methode Ankerfragmente zu kompletten Molekülen zu erweitern, was ein weit verbreitetes Anwendungsszenario in der medizinischen Chemie ist.

Qsearch wurde in verschiedenen Szenarien getestet und hat sich als Generator relevanter Ideen für neue molekulare Grundgerüste bewiesen.

Danksagung

An dieser Stelle möchte ich den Menschen danken, die mich auf dem Weg zum Abschluss dieser Arbeit begleitet haben.

Zuallererst gilt mein Dank meinen Eltern, die während der gesamten Zeit für mich da waren und deren rückhaltloser Unterstützung ich mir stets sicher sein konnte.

An Professor Dr. Matthias Rarey geht mein Dank für die Betreuung und Führung bei der Durchführung der Forschungsarbeiten.

Bei allen neuen und alten Mitarbeitern, mit denen ich in in der AMD-Gruppe gearbeitet habe, möchte ich mich für die gemeinsame Zeit bedanken: Adrian, Andrea, Angela, Axel, Birte, Björn, Christin, Christian, Katrin, Nadine, Ingo, Jochen, Jörg, Lennart, Melanie, Matthias, Patrick und Robert. Dabei möchte ich mich an dieser Stelle insbesondere alle Mitgliedern des NAOMI-Teams dafür danken, dass sie mir und meiner Arbeit einen neuen Motivationsschub gegeben haben, und ihrem Gelingen ganz entscheidend beigetragen haben. Die Arbeit wäre in dieser Form niemals ohne Euch möglich gewesen.

Sascha möchte ich für eine wunderbare Freundschaft und Zusammenarbeit am ZBH danken. Die gemeinsamen Mittagspausen waren für mich immer ein Tageshöhepunkt.

Frank gebührt mein Dank dafür, dass er nicht zugelassen hat, dass ich mich in der Arbeit verliere, und für den notwendigen Anteil Kultur in meinem Leben gesorgt hat. Vielen Dank für die vielen gemeinsamen Abende.

Darüber hinaus möchte ich mich bei der Fa. Hoffman-La Roche für die finanzielle Unterstützung und besonders bei Olivier Roche und Tanja Schulz-Gasch für die befruchtenden Diskussionen und Eingaben bedanken, die dem Projekt zum Erfolg verholfen haben

Inhaltsverzeichnis

1. Einleitung

1.1. Einführung

Als Wirkstoffe werden alle Substanzen bezeichnet, die einen biologischen Effekt in Organismen hervorrufen.[1] Von besonderem Interesse sind hierbei diejenigen Wirkstoffe, deren biologischer Effekt Einfluss auf den Verlauf einer Krankheit hat, und die daher als Arzneistoffe eingesetzt werden können.

Seit dem Entstehen der rationellen Wirkstoffforschung im frühen 20. Jahrhundert hat sich die pharmazeutische Forschung hauptsächlich auf experimentelle Methoden zur Identifikation von neuen Wirkstoffen konzentriert. In den letzten Jahrzehnten wurden parallel dazu computergestützte Verfahren entwickelt, um die medizinischen Chemiker, die auf diesem Feld tätig sind, zu unterstützen.[2] Dabei zeichnen sich die experimentellen und die computergestützten Methoden durch individuelle Stärken und Schwächen aus, weswegen sie in der pharmazeutischen Forschung komplementär verwendet werden.[3] Während experimentelle Methoden im Allgemeinen teurer sind und mehr Zeit benötigen als die virtuellen Methoden, sind diese wegen den Näherungen in den Modellen, die sie verwenden, um die chemischen Gegebenheiten zu beschreiben, weniger exakt.

Die Wirkstoffforschung ist ein iterativer Prozess, der in verschiedene Stufen aufgeteilt ist, wobei auf jeder einzelnen Stufe computergestützte Methoden unterstützend zum Einsatz kommen können. In dem Prozess zeichnen sich die frühen Phasen – in die auch der Gegenstand dieser Forschungsarbeit fällt – unter anderem dadurch aus, dass noch wenig Informationen zu einem konkreten Problemfall vorliegen. Dabei kann es sich zum Beispiel um Strukturformeln von bekannten aktiven Molekülen handeln oder auch um Hinweise, die von vergleichbaren Zielstrukturen stammen. Das Ziel in dieser Phase ist, auf Grundlage der zusammengetragenen Informationen weiteres Wissen, zum Beispiel in Form von biologisch aktiven Molekülen zu gewinnen.

Die computergestützten Verfahren, die in dieser Phase eingesetzt werden, machen sich zunutze, dass ihre Berechnungen nicht mit physikalisch vorliegen Molekülen durchgeführt werden müssen, sondern rein virtuell vorgenommen werden. Dieser Vorteil lässt sich auf zweierlei Weisen ausspielen: Zum einen lassen sich so sehr große Substanzbiblio-

theken kostengünstig auf aktive Einträge überprüfen, und zum anderen ist es so möglich, Vorhersagen über neuartige Moleküle treffen, ohne sie vorher synthetisiert zu haben. Das Generieren neuartiger chemischer Strukturen im virtuellen Wirkstoffentwurf wird *de novo* Design genannt und ist das Kerngebiet der in dieser Dissertation behandelten Thematik,

1.2. Forschungsrahmen

Im Rahmen dieser Dissertation wurden stochastische Verfahren zum Entwurf neuartiger Grundgerüste aus chemischen Fragmenträumen entwickelt. Folgende Punkte waren dabei von herausragendem Interesse und erfuhren eine besondere Beachtung.

- Die Vorteile von Pharmakophoren im Feld des *de novo* Designs sollten ausgenutzt werden. Sie haben sich in der medizinischen Chemie als Standardmethode etabliert, um die chemischen Eigenschaften zu formulieren, die für die *molekulare Erkennung* von kleinen Molekülen durch biologische Strukturen essentiell sind. Sie zeigen besondere Stärken bei *de novo* Designansätzen, da sie von konkreten molekularen Strukturen abstrahieren, und dadurch die generierten Lösungen nicht von vornherein beeinflussen.

- Genauso sollten die vielen Vorteile, die Fragmenträume in der Formulierung von chemischen Unterräumen bieten, genutzt werden. Sie erlauben es, durch einen relativ kleinen Satz an Fragmenten und Verknüpfungsregeln, einen großen chemischen Unterraum zu definieren, der implizit viele Eigenschaften, wie zum Beispiel die der Wirkstoffähnlichkeit, besitzt. Bestehendes Wissen, das auf diesem Feld gesammelt wurde, sollte in die Methode integriert und erweitert werden.

- Für die durchzuführenden Berechnungen boten sich stochastische Verfahren an, da die zu erwartende Energielandschaft sehr rau und steil war, und diese Verfahren sich in vergleichbaren Forschungsarbeiten als erfolgreich bewiesen haben.

- Um eine reibungslose Kommunikation mit den Anwendern bei dem Kooperationspartner zu gewährleisten, sollte das Programm sowohl intuitiv zu bedienen sein als auch Ergebnisse liefern, die sich mit dem Erfahrungswissen der Anwender decken.

1.3. Aufbau der Dissertation

In den folgenden Kapiteln wird die unter dem Namen Qsearch implementierte Methode zum pharmakophorbasierten Durchsuchen von chemischen Fragmenträumen mithilfe von

stochastischen Verfahren präsentiert. Die Ausführungen sind in mehrere Kapitel gegliedert, die aufeinander aufbauen und die durchgeführten Arbeiten erläutern:

Kapitel 2 liefert eine generelle Einführung in die Problemstellung des rationalen Wirkstoffdesigns und stellt die Konzepte aus dem Feld vor, die für das weitere Verständnis dieser Arbeit vonnöten sind. Hierzu gehört insbesondere das Pharmakophor- und das Fragmentraumkonzept.

Kapitel 3 beinhaltet eine Übersicht über das *de novo* Design im Allgemeinen, und eine Vorstellung von Konzepten aus der Literatur, die sich mit ähnlichen Problemstellungen beschäftigt haben.

Kapitel 4 erläutert die Modelle der medizinischen Chemie, die als Grundlage für die durchgeführten Berechnungen verwendet wurden. Dazu gehören vor allen Dingen die Art und Weise mit der die Fragmenträume und die Pharmakophore beschrieben werden.

Kapitel 5 führt eine genaue Beschreibung der verwendeten Methoden auf, mit denen die in Kapitel 4 vorgestellten Konzepte umgesetzt worden sind. Dies beinhaltet eine ausführliche Beschreibung der zum Einsatz gekommenen Algorithmen.

Kapitel 6 zeigt exemplarisch Ergebnisse, die in vier Beispielszenarien erzeugt wurden, und diskutiert ihre Relevanz für die Wirkstoffforschung. Darauf folgt eine Diskussion mit einem besonderen Schwerpunkt auf den Einschränkungen die sich für die Methode ergeben.

Kapitel 7 schließt die Arbeit mit einer Zusammenfassung und einem Ausblick ab.

2. Grundlagen

In diesem Kapitel wird die Rolle von Qsearch im Kontext der Wirkstoffforschung erläutert, sowie verschiedene Schlüsselkonzepte aus der medizinischen Chemie vorgestellt, die für das Verständnis der implementieren Methoden wichtig sind. Im ersten Teil werden im Rahmen einer kurzen Einführung die Interaktionen präsentiert, die zwischen Wirkstoff und Zielstruktur ausgebildet werden, und daraufhin das Pharmakophorkonzept dargestellt, das die Grundlage für die Bewertung von Molekülen durch Qsearch liefert. Anschließend wird das Prinzip der molekularen Grundgerüste eingeführt und der bioisostere Wechsel zwischen ihnen erläutert. Diese Konzepte sind wichtig, um die Ergebnisse in Kapitel 6 einordnen zu können. Letztendlich wird die Technik der Fragmenträume vorgestellt, die als Quelle von neuartigen Grundgerüsten verwendet wird.

2.1. Arbeitsablauf im Wirkstoffdesign

Die Entwicklung von neuen Wirkstoffen ist ein komplexer mehrstufiger Prozess, der oft mit dem Bild einer Pipeline beschrieben wird. Auf der einen Seite der Pipeline stehen als Eingabe diverse Quellen für biochemisches Wissen, und auf der anderen Seite Wirkstoffe als Ausgabe, die nach aufwendigen klinischen Studien als Medikamente Anwendung finden. Der gesamte Prozess ist in Abbildung 2.1.1 aufgezeichnet. Für diese Arbeit ist vor allem der erste Abschnitt relevant, der sich mit der Entwicklung von Vorschlägen für neue Wirkstoffen beschäftigt, und wird daher im folgenden erläutert. Die einzelnen Teilüberschriften beziehen sich dabei auf die o.g. Abbildung.

Identifikation von Zielstrukturen In dieser Phase werden biologische Strukturen identifiziert, die für die Veränderung eines Krankheitsverlaufs relevant sind, und die in den darauf folgenden Schritten durch Wirkstoffe beeinflusst werden sollen. Computergestützte Methoden können unter anderem in *Druggability*-Studien oder auch in der Analyse von systembiologischen Netzwerken eingesetzt werden, um Voraussagen über das Potential einer Zielstruktur zu treffen.

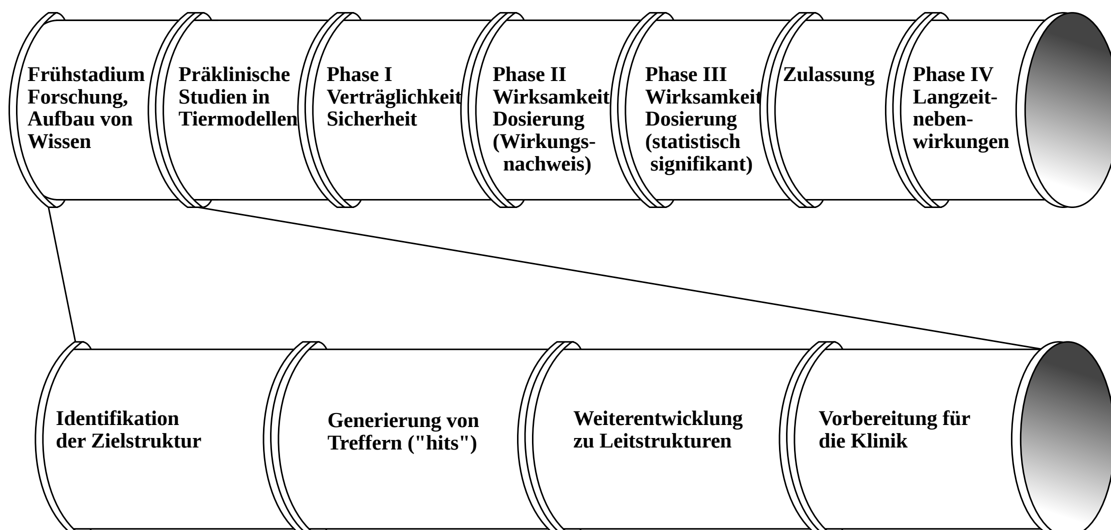


Abbildung 2.1.1.: Pipeline in der Wirkstoffentwicklung. Oben ist die gesamte Pipeline inklusive der klinischen Studien abgebildet, unten ist das Frühstadium der Forschung, in dem neue Substanzen als Wirkstoffkandidaten generiert werden, detailliert dargestellt. Der Arbeitsablauf ist aus [1] übernommen und graphisch aufbereitet worden.

Generierung von Treffern Nachdem eine Zielstruktur identifiziert worden ist, werden in sogenannten *Screening*-Kampagnen systematisch große Mengen an Substanzen experimentell auf eine biologische Aktivität hin überprüft. Eine Substanz, die in dieser Phase gefunden wird, wird als Treffer oder als *hit* bezeichnet. In diesem Feld werden Methoden des *virtuellen Screenings* eingesetzt, um die experimentellen Methoden zu ergänzen. Sie können zum Beispiel dabei helfen, Voraussagen über die potentielle Aktivität eines Moleküls zu treffen. Darüber hinaus spielen computergestützte Methoden eine große Rolle in der Erstellung der chemischen Bibliotheken, die in den Screening-Experimenten eingesetzt werden.

Weiterentwicklung zu Leitstrukturen Sobald *hits* gefunden wurden, wird der chemische Raum, der sie umgibt, extensiv nach Molekülen abgesucht, die sowohl eine gute Aktivität als auch gute pharmakokinetische Eigenschaften besitzen, und die im weiteren Verlauf der Pipeline klinisch untersucht werden. In diesem Feld werden computergestützte Verfahren eingesetzt, um die Natur der chemischen Wechselwirkungen, die der Wirkstoff mit der Zielstruktur eingeht, zu analysieren und zu optimieren. Dabei werden insbesondere bildgebende Verfahren verwendet, die Rückschlüsse auf die ausgebildeten Interaktionen zulassen. Des weiteren finden auch sogenannte *QSAR*-Ansätze („quantita-

tive structure-activity relationship“) Anwendung, die die chemische Struktur der Treffer mit ihrer Aktivität korrelieren und diesbezügliche Vorhersagen ermöglichen.

Positionierung von Qsearch Die in dieser Arbeit vorgestellten Methode zum *de novo* Design ist in dem Arbeitsablauf ungefähr am Ende der Phase „Generieren von Treffern“ bis zum Anfang der Phase „Weiterentwicklung zur Leitstruktur“ anzusiedeln. Sie kann entweder dafür verwendet werden, um auf Grundlage von existierendem Wissen zu einer Zielstruktur neuartige Grundgerüste zu erstellen, die zu Leitstrukturen weiterentwickelt werden, oder sie kann dazu eingesetzt werden, Wechsel in bereits bekannten Grundgerüsten vorzunehmen, um sie mit Hinblick auf bestimmte Parameter zu optimieren.

2.2. Molekulare Interaktionen

Damit ein Wirkstoff seine biologische Wirkung entfalten kann, muss er mit körpereigenen Molekülen interagieren. In der Regel sind dies Makromoleküle, die in Signalkaskaden integriert sind und dadurch große Auswirkungen auf physiologische Prozesse haben. Dabei spielt das Konzept der *molekularen Erkennung* eine herausragende Rolle. Es beschreibt die „Erkennung“ eines Wirkstoffs durch ein Makromolekül, was die Grundlage für eine Einflussnahme ist.

Maßgeblich für die *molekulare Erkennung* in biologischen Systemen sind intermolekulare Interaktionen, wie sie im Feld der *supramolekularen Chemie* untersucht werden. Dabei stehen nicht die kovalenten (also chemischen) Bindungen im Mittelpunkt der Betrachtung, sondern vor allem schwächere Kräfte wie Wasserstoffbrückenbindungen, die *Komplexierung* von Metallionen und die Hydrophobizität. Allerdings sind bisher noch nicht alle an der Erkennung beteiligten Effekte aufgeklärt und definiert worden.[4]

Im folgenden werden die verschiedenen Arten von polaren Interaktionen, die bisher Einzug in das Standardrepertoire der medizinischen Chemie gefunden haben, aufgezählt. Da die Hydrophobizität in Abschnitt 6.5.1 explizit besprochen wird, bleibt sie in dieser Aufzählung außen vor.

Wasserstoffbrückenbindungen Eine der am besten charakterisierten Bindungen in diesem Feld sind die Wasserstoffbrückenbindungen. Sie werden als maßgeblich für große Teile der *molekularen Erkennung* sowie der Stabilität von biologischen Makrokomplexen angesehen.[5]

Die IUPAC definiert sie folgendermaßen:

> Die Wasserstoffbrückenbindung ist eine anziehende Interaktion zwischen einem Wasserstoffatom eines Moleküls oder eines Molekülfragments X-H in dem X elektronegativer als H ist, und einem Atom oder einer Gruppe von Atomen in dem gleichen oder einem anderen Molekül, in der es Hinweise auf eine Bindungsausbildung gibt. (übersetzt aus [6])

Wenn die an einer Wasserstoffbrückenbindung beteiligten Atome in der Form X-H$\cdots$Y-Z dargestellt werden, wird die X-H Komponente als Donor und die Y-Z Komponente als Akzeptor bezeichnet. In der Regel beträgt der Winkel zwischen den Bindungen von X-H und Y 180°, und je näher der Winkel an dieser Idealgeometrie liegt, desto stärker ausgeprägt ist die Interaktion.[7]

Salzbrücken Eine ähnlich gut untersuchte Art von Interaktionen sind die Salzbrücken, die zwischen geladenen funktionellen Gruppen ausgebildet werden.[8] Dabei spielen die elektrostatischen Anziehungskräfte zwischen positiven und negativen Ladungen eine entscheidende Rolle. Prinzipiell gelten für Salzbrücken die gleichen Voraussetzungen wie für Wasserstoffbrückenbindungen, denn auch ihre Ausbildung hängt stark von der geometrischen Ausrichtung der Atome ab. Für die Ausbildung von Salzbrücken werden allerdings noch gegensätzliche Formalladungen an den beteiligten Gruppen benötigt.

Metallkomplexierung Metalle spielen in der strukturellen Stabilität vieler Proteine sowie bei der Vermittlung von Reaktionen in deren aktiven Zentren eine Schlüsselrolle.[9] Hierbei werden Metall-Ionen durch funktionelle Gruppen mit freien Elektronenpaaren innerhalb eines Proteins – und teilweise unter Miteinbeziehung von weiteren Molekülen – *koordiniert*. Wie bei den Wasserstoffbrückenbindungen ist die Geometrie bei Metallkomplexierungen ebenfalls starken Restriktionen unterworfen, was die Position und Orientierung der interagierenden Gruppen angeht.

π – Stacking Das π – Stacking ist ein Spezialfall der London-Kräfte, die die anziehende Kraft zwischen benachbarten, nicht polaren Gruppen durch gegenseitig induzierte Dipole beschrieben. Es ist eine der essentiellen Grundlagen für die Ausbildung von dreidimensionale Struktur von Biomakromolekülen wie DNA.[10] Besonders ausgeprägt sind diese Interaktionen zwischen π–Systemen, die in bestimmten Winkeln zueinander stehen.[11]

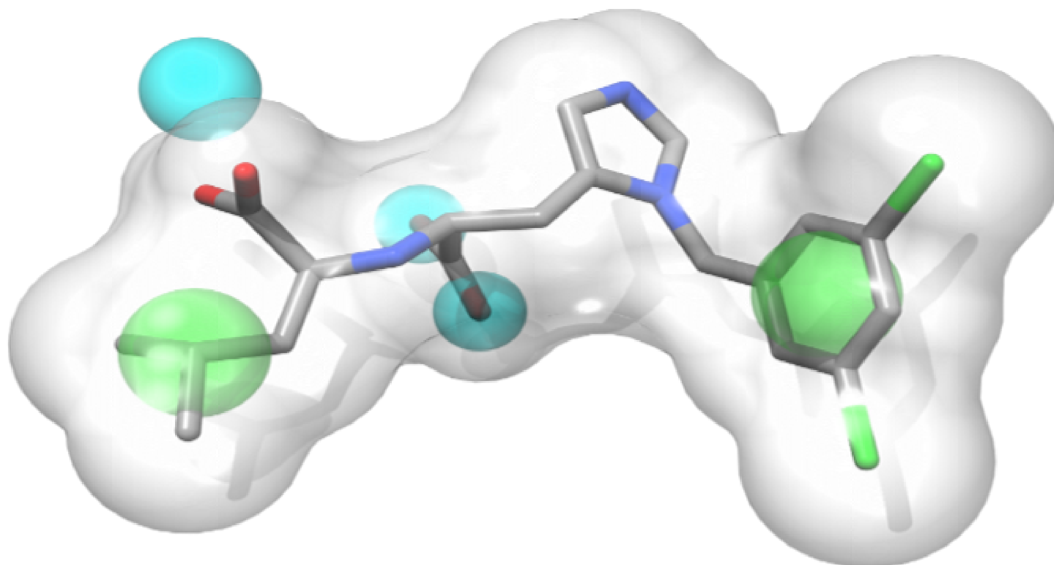


Abbildung 2.3.1.: ACE-2 Pharmakophor nach Rella et al.[12] In grün sind hydrophobe Merkmale dargestellt und in cyan links oben die Richtung eines Akzeptormerkmals und in der Mitte das Zentrum von zwei Akzeptormerkmalen. Die sterischen Eigenschaften sind als Wolke um das Molekül angedeutet.

2.3. Pharmakophore

Pharmakophore sind eine Methode aus der medizinischen Chemie, um Schlüsselinteraktionen zu formulieren, die für die biologische Aktivität von Wirkstoffen relevant sind. Historisch wurde das Konzept 1967 das erste Mal von Lemont B. Kier [13] beschrieben und 1971 als Begriff geprägt[14].[15] Der große Erfolg und die weite Verbreitung des Konzeptes hat die IUPAC dazu bewegt, einige Jahrzehnte später eine Definition für den Begriff zu veröffentlichen:

> Ein Pharmakophor ist die Gesamtheit aller sterischen und elektronischen Eigenschaften die gegeben sein müssen, um eine optimale supramolekulare Interaktion mit einer biologischen Zielstruktur zu ermöglichen und ihre biologische Reaktion auszulösen oder zu blockieren. (übersetzt aus [16])

Begleitend zu dieser Definition wird in einer Ergänzung ausdrücklich auf die korrekte Verwendung hingewiesen, um der häufigen falschen Benutzung des Begriffs in der medizinischen Chemie entgegenzuwirken:

> Ein Pharmakophor repräsentiert kein konkretes Molekül und keine explizite Anordnung von funktionellen Gruppen, sondern ist eine rein abstraktes Kon-

> zept, das die Fähigkeit verschiedener Stoffklassen beschreibt, Interaktionen mit ihrer Zielstruktur auszubilden. Der Pharmakophor kann als der größte gemeinsame Nenner, auf den eine Menge an aktiven Molekülen gebracht werden kann, betrachtet werden. Diese Definition ersetzt die weit verbreitete falsche Benutzung des Begriffes in der Literatur der medizinischen Chemie, die den Begriff für einfache chemische Funktionalitäten wie Guanidine, Sulfonamide oder Dihydroimidazole, oder für typische strukturelle Skelette wie Flavone, Phenothiazine, Prostaglandine oder Steroide verwendet. (übersetzt aus [16])

Wie aus der Definition ersichtlich, lassen sich die Eigenschaften in sterische Eigenschaften und elektronische Merkmale aufteilen. Die sterischen Eigenschaften beschreiben den Raum, den ein ein Molekül innerhalb der Zielstruktur einnehmen kann, ohne mit ihr zu kollidieren.[1] Dieser Ort innerhalb der Zielstruktur entspricht in der Regel dem aktiven Zentrum eines Enzyms, kann aber auch an anderen Stellen, wie zum Beispiel den Erkennungsstellen für Protein-Protein Interaktionen liegen. Um diesen Raum für eine konkrete Zielstruktur zu bestimmen, gibt es zwei verschiedene Möglichkeiten. Wenn dreidimensionale Informationen über die Zielstruktur vorliegen, so kann das aktive Zentrum direkt als Matrize verwendet werden, um das zur Verfügung stehende Volumen zu definieren. Liegen keine solche Informationen vor, so wird dieser Raum von Molekülen abgeleitet, die bekanntermaßen an dieser Stelle binden. Dazu werden ihre dreidimensionalen Strukturen entweder manuell oder algorithmisch überlagert und als Basis für eine Hypothese der Form des aktiven Zentrums verwendet.

Die elektronischen Merkmale beschreiben die chemischen Eigenschaften das Moleküls, die idealerweise komplementär zu denen der Zielstruktur sind. Um diese zu erfüllen, müssen sie dergestalt sein, dass Schlüsselinteraktionen zwischen dem Wirkstoff und der Zielstruktur ermöglicht werden. Diese Informationen können hierbei analog zu den sterischen Eigenschaften entweder über die dreidimensionale Struktur des aktiven Zentrums oder durch die Überlagerung von bekannten aktiven Molekülen abgeleitet werden. In der Regel werden die in Abschnitt 2.2 aufgeführten Interaktionen in verwendet, um ein Pharmakophormodell zu beschreiben.[18]

[1] Der Term „kollidieren“ ist eine bildliche Beschreibung der tatsächlichen quantenmechanischen Prozesse. Nach dem Pauli-Prinzip[17] ist es zwei Elektronen mit gleicher Energie nicht möglich, am gleichen Ort mit dem gleichen Spin zu existieren. Dies resultiert in einer Abstoßung der Elektronen, und somit ihrer dazugehörigen Atome.

2.4. Molekulare Grundgerüste und Grundgerüstwechsel

2.4.1. Molekulare Grundgerüste

Wirkstoffe lassen sich aufgrund ihrer chemischen Struktur in Klassen mit gleichen Grundgerüsten einordnen, wobei sich für diese Einordnung noch kein fester Satz an Regeln etabliert hat. Teilweise wird sie über die zentralen Strukturelemente durchgeführt, teilweise über spezifische funktionelle Gruppen in den Molekülen und teilweise über eine Kombination der beiden Ansätze.[19] Die Komplexität des Themas zeigt sich unter anderem darin, dass von der IUPAC bisher keine allgemein verbindliche Definition des Begriffs veröffentlicht wurde.

Das Konzept der Grundgerüste selbst ist in der Chemie schon lange bekannt, wurde aber vor 1969 nicht explizit formuliert. (vgl. ebd.) In diesem Jahr fanden Reich und Cram[20] folgende Formulierung:

> Das Ringsystem [in diesen Molekülen] ist hochgradig starr, und kann als Grundgerüst dienen, um die funktionellen Gruppen geometrisch gegeneinander so auszurichten, dass systematische Untersuchungen des Einflusses von Ringschlüssen und funktionellen Gruppen auf die physikalischen und chemischen Eigenschaften möglich werden. (übersetzt aus [20])

Diese einfache Definition beschreibt das Konzept des Grundgerüsts: ein über verschiedene Moleküle invarianter Teil, der funktionelle Gruppen geometrisch zueinander positioniert. Allerdings lässt sie eine formale Definition der Elemente, die zu einem Grundgerüst gezählt werden müssen und welche nicht, vermissen. Eine solche Definition wurde erst in den späten neunziger Jahren von Bemis und Murcko[21] angegangen. Sie basiert auf der Reduktion der Komplexität von Molekülen durch Nichtbeachtung von Element- und Bindungsinformationen sowie dem Verwerfen von Seitenketten. In dieser Definition werden Atome in Klassen eingeteilt, aus denen sich das Grundgerüst erstellen lässt:

Ringsysteme Die Zweifach-Zusammenhangskomponenten eines Moleküls, wenn es als molekularer Graph betrachtet wird.

Verbindungsatome Die Atome, die auf einem Pfad zwischen zwei Ringsystemen liegen und zu keinem Ringsystem gehören.

Seitenketten Die Atome, die weder in Ringsystemen noch Verbindungsatome sind.

Grundgerüst Die Vereinigung aller Atome in einem Molekül, die entweder einem Ringsystem angehören oder als Verbindungsatom fungieren.

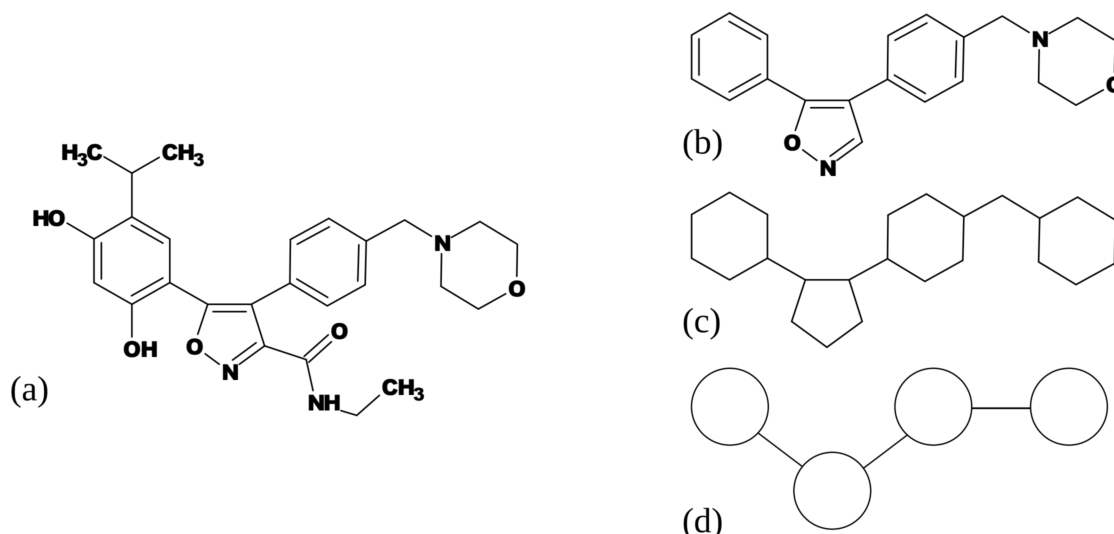


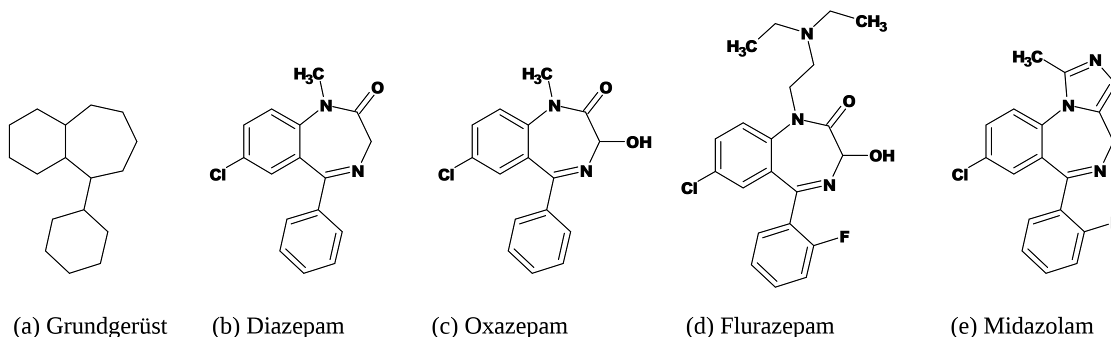
Abbildung 2.4.1.: Der HSP90 Inhibitor NVP-AUY922 (a), und das molekulare (b), graphbasierte (c), und reduzierte Grundgerüst (d). (nach Langdon et al.[19])

Diese Definition kommt der intuitiven Definition von Grundgerüsten in der medizinischen Chemie nahe, ist allerdings nicht vollständig. Als Beispiele seien die Benzodiazepine und die Chinolone in den Abbildungen 2.4.2 beziehungsweise 2.4.3 aufgeführt, die im folgenden erläutert werden:

Bei den Benzodiazepinen[2] überschneidet sich das graphbasierte Grundgerüst mit dem in der medizinischen Chemie üblichen Gerüst. Allerdings beinhaltet dieses zusätzlich noch Informationen über die Art der Bindungen, die Elemente der Atome, und die Aromatizität der Ringe. Außerdem wird deutlich, dass die graphbasierte Definition nicht den Fall beachtet, in dem ein Grundgerüst als Substruktur in einem anderen enthalten ist. Dies wird in der in Abbildung 2.4.2(e) abgebildeten Struktur klar. Sie zählt nach der klassischen Definition mit zu den Benzodiazepinen, würde allerdings Bemis und Murcko in eine eigene Klasse fallen, da der anellierte Ring zu dem Grundgerüst hinzugerechnet wird.

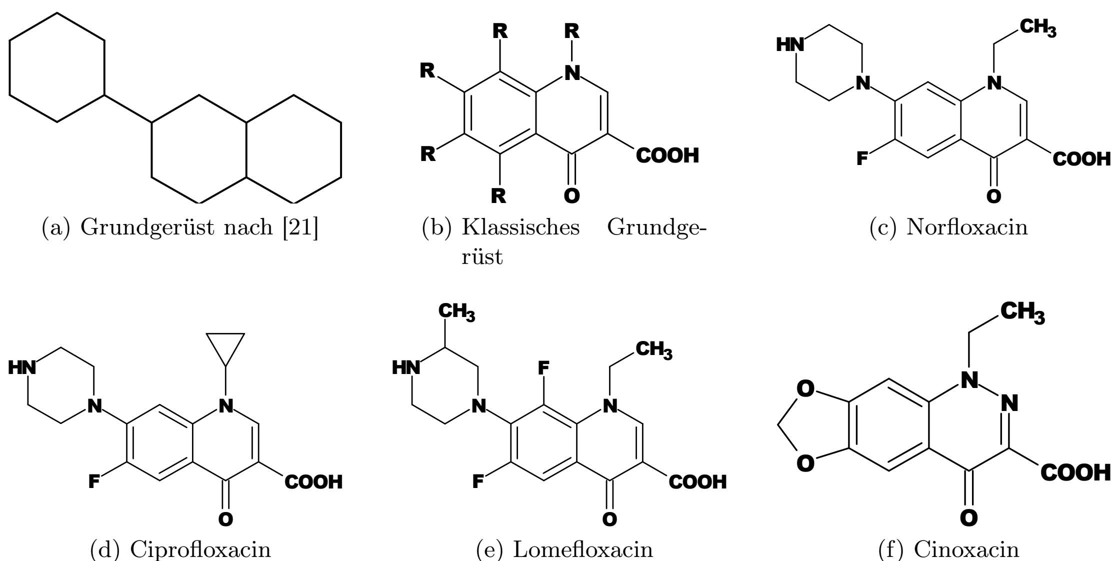
Um die Schwierigkeiten bei der Definition von Grundgerüsten weiter zu erläutern, seien die Chinolone aufgeführt, die als Antibiotika breite Anwendung finden. Das in der medizinischen Chemie verwendete Grundgerüst ist in Abbildung 2.4.3 aufgeführt und besteht aus einem Chinolin-Doppelring mit einer Carbonsäure in 3-Position und einem Carbonyl in 4-Position. Das Grundgerüst nach Bemis und Murcko würde diese

[2] Benzodiazepine sind eine Klasse von kompetitiven GABA-Rezeptorinhibitoren, die das gleiche Grundgerüst teilen und je nach Dekoration angstlösend, muskelrelaxierend, sedierend und/oder hypnotisch wirken.



(a) Grundgerüst (b) Diazepam (c) Oxazepam (d) Flurazepam (e) Midazolam

Abbildung 2.4.2.: Verschiedene Benzodiazepine und das Grundgerüst (a), das fast alle Mitglieder der Klasse teilen. Die unter (f) aufgeführte Struktur hat das Grundgerüst als Substruktur, würde allerdings nach der Definition von Bemis und Murcko[21] nicht zu der gleichen Klasse gehören, da der Siebenring eine Bindung mit einem weiteren Fünfring teilt.



(a) Grundgerüst nach [21] (b) Klassisches Grundgerüst (c) Norfloxacin

(d) Ciprofloxacin (e) Lomefloxacin (f) Cinoxacin

Abbildung 2.4.3.: Verschiedene Chinolone (c-e) und zwei Grundgerüste (a,b). Mit (f) ist ein Chinolon aufgeführt, das der Grundstruktur in (a) nicht entspricht.

Information nicht korrekt widerspiegeln und stattdessen in dem in Abbildung 2.4.3(a) gezeigten resultieren. Die Information über funktionellen Gruppen, die für diese Klasse charakteristisch sind, wäre demnach genauso wenig wie die Aromatizität enthalten.

Da eine abschließende Klärung des Grundgerüstbegriffs noch ungeklärt ist, wird in dieser Arbeit je nach Anwendung sowohl auf die graphbasierte als auch auf die tradierte Grundgerüstdefinitionen zurückgegriffen, um die Ergebnisse von Qsearch zu analysieren.

2.4.2. Grundgerüstwechsel

Unter einem Grundgerüstwechsel versteht man in der pharmazeutischen Forschung die Veränderung des chemischen Grundgerüsts[3] eines Wirkstoffs unter Beibehaltung der biologischen Aktivität. Grundgerüstwechsel sind häufig das Ziel von pharmazeutischen Forschungskampagnen, und die Arbeiten an Qsearch wurden unter anderem dafür durchgeführt, medizinische Chemiker hierbei zu unterstützen. Ein Grundgerüstwechsel kann durch eine Reihe von Gründen notwendig werden. Dazu gehört vor allem die Verbesserung von pharmakologischen Eigenschaften wie die Wasserlöslichkeit, Potenz oder Toxizität. Darüber hinaus lassen sich mit Grundgerüstwechseln auch Patente von Mitbewerbern umgehen und neues geistiges Eigentum schaffen. Eine Reihe von veröffentlichten Artikeln ([22, 23, 24]) bietet eine gute Übersicht über die aktuellen Methoden und Anwendungen auf diesem Feld.

Die Unsicherheiten bei der Definition eines chemischen Grundgerüsts ziehen sich weiter durch bis zum Grundgerüstwechsel. Da für ersteres keine schlüssige Definition existiert, kann letzteres offensichtlich auch nicht befriedigend beschrieben werden. Der Übergang von einfachen Modifikationen eines Grundgerüsts zu „Sprüngen“ ist fließend und nicht ausreichend definiert. Um Mehrdeutigkeiten vorzubeugen werden in diesem Kontext nur Veränderungen im Grundgerüst als Grundgerüstwechsel angesehen, durch die die Topologie des Grundgerüstes sich verändert (vgl. Abbildung 2.4.4).

2.5. Chemische Fragmenträume

Chemische Fragmenträume werden in der computergestützten medizinischen Chemie dazu verwendet, (zum Teil sehr große) Mengen von Molekülen in einer kompakten Art und Weise zu beschreiben. Dabei finden Fragmenträume nicht nur im *de novo* Design Anwendung, sondern auch bei der Erstellung von kombinatorischen Bibliotheken, wobei sich je-

[3] Wie im vorherigen Abschnitt erklärt, ist die Definition der Grundgerüste im Allgemeinen noch nicht geklärt. Nichtsdestotrotz gibt es für praktisch alle Wirkstoffklassen in der medizinischen Chemie tradierte Grundgerüstkonzepte, die den jeweiligen Ausführungen zugrunde gelegt werden.

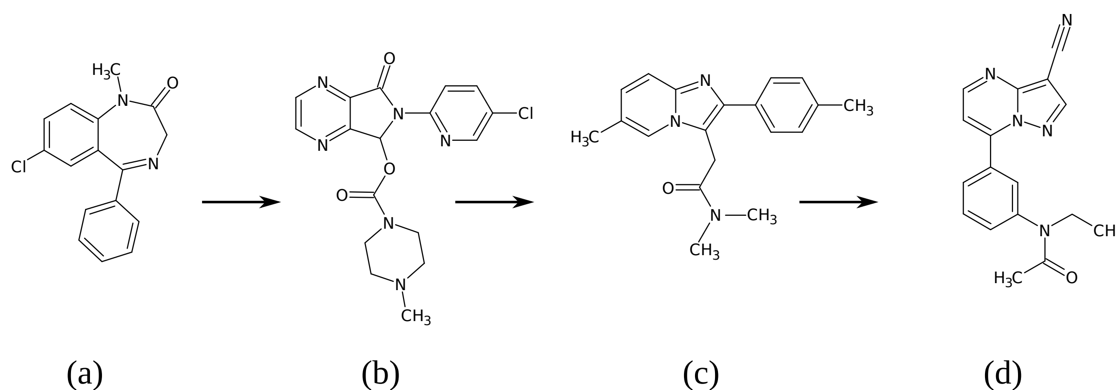


Abbildung 2.4.4.: Beispiel eines Grundgerüstwechsel aus [22]. Ausgehend von Diazepam ((a), Benzodiazepine) wurde das Grundgerüst über Zopiclon ((b), Zyklopyrrole), und Zolpidem ((c), Imidazopyridine), zu Zaleplon ((d), Pyrazolpyrimidine) gewechselt.

weils die Zusammenstellung der Fragmenträume wegen der verschiedenen Zielsetzungen unterscheidet. Fragmenträume, die für die Abbildung von kombinatorischen Bibliotheken erzeugt worden sind, modellieren in der Regel die Dekoration von zentralen Kernfragmenten mit einem Satz an R-Gruppen.[25] Fragmenträume hingegen, die für die Verwendung im *de novo* Design erstellt worden sind, sind in der Regel in ihrer Zusammenstellung weitaus diverser, da sie darauf optimiert sind, möglichst verschiedenartige Strukturen zu erzeugen, die darüber hinaus wirkstoffähnliche Eigenschaften besitzen. [26] Eine ausführlichere Übersicht über die verschiedenen Fragmentraumtypen und ihre Unterschiede findet sich in Abschnitt 6.6. In den folgenden Ausführungen werden ausdrücklich nur Fragmenträume besprochen, die für das *de novo* Design Anwendung entworfen worden sind.

Fragmenträume wurden ursprünglich von Lewell et al. entwickelt, um klassische Syntheseprotokolle der kombinatorischen Chemie in Computermodellen abzubilden.[27] Dazu wurde unter dem Namen RECAP ein Satz an Regeln entwickelt, nach dem Moleküle an definierten Punkten in molekulare Fragmente aufgetrennt und wieder zusammengefügt werden können, um neue Moleküle zu erzeugen. Das Konzept der chemischen Fragmenträumen wurde in vielen *de novo* Computerprogrammen verwendet (vgl.[28, 29, 30, 31]) und von verschiedenen Gruppen weitergeführt und verfeinert ([26, 32]).

Ein chemischer Fragmentraum besteht aus einer Menge von chemischen Fragmenten mit definierten Verknüpfungspunkten und einem Satz an Regeln, nach denen diese zusammengefügt werden können. Implizit wird so in einer kompakten Form eine Menge an möglichen Molekülstrukturen definiert, die durch das Verknüpfen der Fragmente nach

den dazugehörigen Regeln entstehen können. Eine entsprechende Auswahl der Verknüpfungsregeln kann im Vergleich zu atombasierten Verfahren die Synthetisierbarkeit der erzeugten Moleküle erhöhen und die Zusammenstellung der Fragmente die Wirkstoffähnlichkeit verbessern.[32] Die meisten Methoden verwenden Fragmenträume nur als kompakte Beschreibung von chemischen Räumen, die sie in der Regel enumerieren oder mithilfe von stochastischen Algorithmen durchsuchen. Algorithmisch zieht nur FTrees-FS[29] Vorteile aus der kombinatorischen Natur der Fragmenträume.

Die Nützlichkeit von chemischen Fragmenträumen in der computergestützten Wirkstoffforschung wurde in vielen Studien gezeigt. Sie werden häufig im *de novo* Design eingesetzt und haben den atombasierten Ansatz praktisch abgelöst.[33] Eine aktuelle Übersicht über aktuelle Anwendungsbeispiele von Fragmenträumen kann zum Beispiel in [34] gefunden werden.

3. *de novo* Design

In diesem Kapitel wird das Feld des *de novo* Designs in der virtuellen Wirkstoffforschung behandelt. Nach einer Einführung in die Thematik und einer Übersicht wird ein Einblick in den aktuellen Stand der Forschung gegeben. Dabei werden sowohl rezeptor- als auch ligandbasierte Ansätze präsentiert, da Qsearch Konzepte aus beiden Feldern verwendet.

3.1. Einführung und Überblick

de novo Design ist die Disziplin der virtuellen Wirkstoffentwicklung in der computergesteuert neuartige Strukturen von Molekülen nach gewünschten Kriterien erzeugt werden. Die eingesetzten Techniken lassen sich grob zwei Anwendungsfeldern zuordnen. Zum einen können sie zum Generieren von gänzlich neuartigen Grundgerüsten verwendet werden, die dann von *Hits* zu sogenannten *Leads* weiterentwickelt werden, und zum anderen können sie dazu eingesetzt werden, Strukturänderungen an bereits bekannten Grundgerüsten vorzunehmen. Dabei wird in der Regel versucht, strukturelle Motive zu ersetzen, die unerwünscht sind, weil sie entweder negative physiologische Eigenschaften haben, oder bereits als geistiges Eigentum geschützt sind. Die IUPAC definiert das *de novo* Design folgendermaßen:

> *de novo* Design ist das Design von bioaktiven Verbindungen durch das inkrementelle Aufbauen innerhalb eines Rezeptormodells oder eines aktiven Zentrums eines Enzyms, dessen Struktur durch Röntgenstrukturanalyse oder Kernspinresonanzspektroskopie aufgeklärt wurde. (übersetzt aus [16])

Die Festlegung auf den inkrementellen Aufbau ist dabei klar den zu Zeiten der Definitionserstellung (1998) bekannten Methoden geschuldet, die sich aller dieser Technik bedient haben. Die methodische Pionierarbeit wurde dabei von Programmen wie LUDI[35], SPROUT[36] und PRO-LIGAND[37] geleistet. Inzwischen haben weitere Methoden wie die erschöpfende Analyse aller Moleküle eines Fragmentraums[38] oder die Rekombination von bekannten aktiven Substanzen[39] Einzug in das Repertoire der Methoden gehalten. Hervorzuheben ist darüber hinaus die Fragmentersetzung, bei der zentrale Komponenten

von Molekülen automatisiert ersetzt werden (z.B. in Recore[40]), und die ebenfalls dem Feld des *de novo* Designs zuzuordnen ist.

3.2. Der chemische Raum

Die Anzahl an theoretisch möglichen Molekülen ist enorm. Selbst wenn nur Moleküle betrachtet werden, die aus den organischen Elementen bestehen[1] und chemisch stabil sind, bringt deren Aufzählung moderne computergestützte Methoden an ihre Grenzen. Das zur Zeit anspruchsvollste Projekt, die GDB-13, zählt genau diese Moleküle bis zu einer Größe von 13 Atomen auf und kam nach knapp 40.000 Stunden Rechenzeit auf knapp unter 10^9 Verbindungen.[41] Die Anzahl der möglichen Moleküle mit einer Größe bis 30 Atomen – in diesem Fall ohne die Berücksichtigung von Chlor – steigert sich aufgrund der kombinatorischen Explosion auf geschätzte 10^{60}.[42] Erschöpfende Berechnungen auf diesen Bibliotheken verbieten sich allein schon durch den zeitlichen Aufwand, den sie verursachen würden. Daher werden verschiedene Techniken eingesetzt um den Raum auf relevante Bereiche zu beschränken.

Als vorteilhaft hat sich hierbei die Verwendung von chemischen Fragmenträumen (vgl. Abschnitt 2.5) erwiesen.[34] Sie beschränken den Suchraum auf Moleküle, die sich durch die Verknüpfung von vorgegebenen Fragmenten erstellen lassen. Bei einer intelligenten Fragmentauswahl lässt sich so der Raum auf einen Teilbereich reduzieren, dessen Moleküle durchschnittlich wirkstoffähnlicher und leichter synthetisierbar sind, als bei einer rein zufälligen Auswahl.[32, 43]

Alternativ zur Reduktion des Suchraums kann auch die Fragestellung eingeschränkt werden. Anstatt komplette Moleküle zu konstruieren, wird so zum Beispiel nur ein zentrales Strukturelement nach bestimmten Kriterien ausgetauscht, wobei die Seitenketten fixiert bleiben.[24] Bis vor kurzem wurde dieser Ansatz vor allen Dingen durch die manuellen Kombination und Interpretation verschiedener Methoden umgesetzt[23], bekam aber durch den Einsatz von geometrisch indizierten Fragmentdatenbanken[40] einen neuen Schub. Genauso kann der umgekehrte Ansatz verfolgt werden, und ein zentrales Grundgerüst so mit Seitenketten dekoriert werden, dass möglichst gute Interaktionen mit der Zielstruktur ausgebildet werden.[44] Dieser Ansatz findet insbesondere in Verbindung mit den Methoden der kombinatorischen Chemie im Bibliotheksdesign Anwendung.[45]

[1] Als organisch werden in der GDB-13 alle Moleküle betrachtet, die aus Kohlenstoff, Stickstoff, Sauerstoff, Schwefel und Chlor bestehen.

3.3. Rezeptorbasierte *de novo* Methoden

Die ersten publizierten Methoden auf dem Feld des *de novo* Designs basieren auf einer dreidimensionalen Aufklärung der Zielstruktur[2] und konstruieren neue Moleküle, indem sie diese in der Zielstruktur entweder atom- oder fragmentbasiert inkrementell aufbauen.[46] Qsearch teilt mit diesen Programmen vor allem das Problem, chemisch valide Strukturen zu generieren. Daher sind hier exemplarisch mehrere Programme aufgeführt, um verschiedene Ansätze darzustellen, die zur Lösung des Problems veröffentlicht worden sind. Dabei erfahren die Aspekte, die Teile der in Qsearch behandelten Problemstellung betreffen, besondere Beachtung. Ein weitergehende Darstellung aktueller *de novo* Designmethoden lässt sich in einer Reihe von Übersichtsartikeln der letzten Jahre[24, 34, 46, 47, 48] gewinnen.

3.3.1. LUDI

Eines der ersten *de novo* Programme wurde von Böhm 1992 unter dem Namen LUDI[35] veröffentlicht. Das Programm läuft nach dem sogenannten place-and-link Verfahren ab. In einem ersten Schritt werden Stellen im aktiven Zentrum identifiziert, die für bestimmte Schlüsselinteraktionen verantwortlich sind. Dort werden Startfragmente platziert, die zu diesen Interaktionen kompatibel sind, und in einem weiteren Schritt zu vollständigen Molekülen verbunden. LUDI verwendet sowohl für die Startfragmente als auch für die Verbindungsfragmente eine fest definierte Bibliothek, die durch Benutzer erweitert werden kann. Allerdings basiert die Verknüpfung der Fragmente nicht auf einem retrosynthetischen Regelwerk wie den RECAP Regeln[27], sondern auf einer Verbindung an freien Wasserstoffatomen.

Abgesehen von der historischen Bedeutung von LUDI werden viele Elemente des Interaktionsmodells, das in dem Programm zum Einsatz gekommen ist, in Qsearch aufgegriffen.

3.3.2. FlexNovo

Das Programm FlexNovo[31] basiert auf dem inkrementellen Aufbau von Molekülen des FlexX-Algorithmus[49]. Als Eingabe dient hier ebenso wie bei LUDI die dreidimensionale Struktur des aktiven Zentrums. Dafür wird ein k-greedy Verfahren, das als Tiefensuche gestaltet ist, verwendet um die Liganden im aktiven Zentrum inkrementell aufzubauen.

[2] Der Name der rezeptorbasierten Methoden leitet sich daher, dass die Zielstruktur häufig (aber nicht immer) ein biologischer Rezeptor ist.

Der Algorithmus startet mit der automatisierten Platzierung von Fragmenten in dem aktiven Zentrum[50], wobei versucht wird, möglichst gute Interaktionen zwischen dem Fragment und der Zielstruktur auszubilden. Von allen platzierten Fragmenten werden die k besten für die weiteren Berechnungen berücksichtigt. In den folgenden Schritten werden diese Fragmente zu vollständigen Molekülen ausgebaut, indem an jedes ihrer Linkatom alle kompatiblen Fragmente des Fragmentraums angebaut werden. Von diesen Teillösungen werden nun erneut die k besten ausgesucht und für die nächste Stufe des inkrementellen Aufbaus berücksichtigt. Sobald ein Molekül bestimmte vorgegebene Kriterien erfüllt, so wird es als Lösung ausgegeben.

Qsearch bedient sich des Fragmentraumkonzepts, das FlexNovo verwendet um den Suchraum zu beschreiben, wobei die konkreten Unterschiede und Gemeinsamkeiten in Abschnitt 4.1 eingehend besprochen werden.

3.3.3. SkelGen

Unter dem Namen SkelGen wurde von der Firma De Novo Pharmaceuticals eine Reihe von Methoden veröffentlicht, die sich mit der Problemstellung des *de novo* Designs beschäftigen. Eine genaue Aufschlüsselung der einzelnen verwendeten Methoden gestaltet sich schwierig, da aufgrund der kommerziellen Verwertbarkeit der Forschungsergebnisse nur wenige Details veröffentlicht wurden. Stattdessen finden sich in der wissenschaftlichen Literatur Veröffentlichungen, in denen von der erfolgreiche Anwendung des Programms berichtet wird, und die teilweise Details zu den zum Einsatz gekommenen Methoden beinhalten. Im folgenden ist auf Grundlage dieser publizierten Artikel eine Übersicht über die zur Anwendung gekommenen Methoden zusammengestellt.[3]

Die Grundlage für SkelGen liefern zwei Algorithmen, die Ende der neunziger Jahre veröffentlicht wurden. In den dazugehörigen Artikeln wird beschrieben, wie neuartige Kohlenstoffgrundgerüste in dem aktiven Zentrum eines Proteins konstruiert[30] und anschließend dekoriert[51] werden. Die Grundgerüste werden durch das Verknüpfen von undekorierten Kohlenstofffragmenten generiert, wobei sie an allen terminalen Atomen verknüpft werden können. Jedes Grundgerüst, das so konstruiert wurde, wird nachträglich so mit Heteroatomen dekoriert, dass es optimale Interaktionen in dem aktiven Zentrum ausbildet.

Den Kern des Algorithmus zur Erstellung neuer Grundgerüste in SkelGen stellt eine Monte-Carlo-basierte Methode. Dabei setzt sich der Suchraum aus allen möglichen Kohlenstoffgerüsten und den möglichen Orientierungen und Konformationen dieser Gerüste

[3] Die Firma De Novo Pharmaceuticals hat inzwischen ihre Geschäftstätigkeit eingestellt. Daher lassen sich weder die publizierten Ergebnisse noch die Methode exakt rekonstruieren.

zusammen. Ein neuer Zustand in der Optimierung wird in einem fest vorgegebenen Verhältnis entweder durch die Veränderung der dreidimensionalen Koordinaten des aktuellen Grundgerüsts oder eine Veränderung des Grundgerüsts selbst erzeugt. Die Bewertung der einzelnen Moleküle geschieht durch eine komplexe Bewertungsfunktion, die nicht nur die mit dem Protein ausgebildeten Interaktionen berücksichtigt, sondern auch molekulare Deskriptoren wie das Molekulargewicht, die Hydrophobizität, Pharmakophormerkmale, und die Anzahl an Ringsysteme miteinbezieht.[4] Die Bewertungsfunktion kann auch dahingehend erweitert werden, dass sie in ligandbasierten Szenarien eingesetzt werden kann[52], und Terme für die Rezeptorflexibilität mit berücksichtigt[53].

Die generelle Anwendbarkeit von SkelGen wird in einer Reihe von Publikationen beschrieben. Hierbei ist eine Validierungsstudie von Stahl et al.[54] herauszuheben, in der die Ergebnisse von SkelGen retrospektiv mit bekannten Liganden verglichen werden. In ihr wird gezeigt, wie SkelGen für eine Reihe von Zielstrukturen (Estrogen Rezeptor, COX-2, Stromelysin, und CDK-2) eine Anzahl an bekannten Grundgerüsten reproduzieren konnte, und neuartige Strukturen generiert hat, die von medizinischen Chemikern als relevant angesehen wurden. Unter den Lösungen fanden sich auch Moleküle mit neuartigen Grundgerüsten, die den Estrogen-Rezeptor inhibieren, was durch Experimente bestätigt werden konnte.[55]

3.4. Ligandbasierte *de novo* Methoden

Qsearch gehört zu der Klasse der ligandbasierten *de novo* Methoden. Methoden dieser Klasse basieren auf dem *active analog* Prinzip, das die biologischen Aktivität von Molekülen auf einen Satz an Schlüsselinteraktionen zurückführt.[56] Auf Basis dieses Prinzips wurden verschiedene Methoden entwickelt, die die Ähnlichkeit von Molekülen in Bezug auf die (Schlüssel-)Interaktionen die sie ausbilden, bewerten. Dabei werden die ausgebildeten Interaktionen nicht direkt miteinander verglichen, sondern nur indirekt über verschiedene molekulare Deskriptoren.

Die Deskriptoren, die dabei zum Einsatz kommen, können entweder auf zwei- oder dreidimensionalen Beschreibungen der Moleküle basieren. Erstere verwenden zum Beispiel die Topologie der Moleküle[57] oder sogenannte *Fingerprint*-Methoden[58], um Aussagen über die Ähnlichkeit zu bekannten aktiven Molekülen zu machen. Letztere vergleichen Moleküle über ihre räumliche Ausdehnung und die Anordnung ihrer elektronischen Eigenschaften.[59] Da beide Ähnlichkeitsmaße auf dem Vergleich zu bekannten Liganden basieren, werden sie *ligandbasiert* genannt. Im folgenden werden verschiedene ligandba-

[4] Eine ausführliche Liste ist in [30] zu finden.

sierte *de novo* Ansätze präsentiert.

3.4.1. Flux

Das Programm Flux[60] verwendet einen populationsbasierten evolutionären Algorithmus um durch den chemischen Raum zu navigieren. Dafür werden evolutionäre Mechanismen auf virtuelle „Molekülpopulationen“ angewendet, und ein „Selektionsdruck“ für eine möglichst hohe Ähnlichkeit zu einer Referenzstruktur aufgebaut. Die Fitnessfunktion, die dabei zum Einsatz kommt, setzt sich aus zwei molekularen Fingerprints zusammen. Einerseits einem, der auf den Ghose-Crippen Atomtypen[61] basiert[62], und andererseits auf dem Daylight Fingerprint[5]. Um neue Moleküle zu erzeugen, verknüpft das Programm Fragmente der COBRA-Bibliothek[43]. Eine Besonderheit ist hierbei, dass nicht nur einfache Mutationen der einzelnen Moleküle möglich sind, sondern auch sogenannte *Crossover* Operationen, bei denen zwischen den Individuen innerhalb einer Population Informationen ausgetauscht werden.[63]

Mithilfe von Flux wurden unter anderem neuartige Moleküle mit Aktivität gegen HI-Viren vorhergesagt, die auch experimentell validiert wurde.[64] Des weiteren konnte mit dem Programm auch eine neue Leitstruktur für das Blockieren von Kalium-Kanälen vorhergesagt werden, die ebenfalls *in vitro* bestätigt wurde.[28]

3.4.2. NovoFLAP

In Kooperation mit der Firma Tripos hat die Firma AstraZeneca das Programm NovoFLAP entwickelt. Es liefert offenbar gute Ergebnisse[65], befindet sich aber nur im internen Einsatz, weswegen die verwendete Methodik nicht im Detail beschrieben wurde. NovoFLAP kombiniert die Anwendung eines evolutionären Algorithmus um Fragmenträume zu durchsuchen mit einer flexiblen Ligandausrichtung um die Fitness der einzelnen Teillösungen zu bestimmen. Die Suche in dem Fragmentraum wird mithilfe des Programms EA-Inventor[6] durchgeführt, und die Bewertung der Teillösungen mithilfe des Programms FLAP[7].

Um ein Molekül zu bewerten, das im Verlauf der Optimierung erzeugt wurde, verwendet FLAP einen zweiphasigen Algorithmus. In der ersten Phase kommen die externen Programme Omega[8] und ROCS[9] zum Einsatz. Für jedes erzeugte Molekül wird durch

[5] Daylight Chemical Information Systems Inc.: Irvine, CA (USA). `www.daylight.com`
[6] EA-Inventor; Tripos L.P.: St. Louis, MO (USA).`www.tripos.com`
[7] FLAP („flexible ligand alignment protocol“); AstraZeneca Pharmaceuticals: Wilmington, DE (USA). Nicht veröffentlicht.
[8] Omega; OpenEye Scientific Software Inc.: Santa Fe, NM (USA). `www.eyesopen.com/omega`
[9] ROCS; OpenEye Scientific Software Inc.: Santa Fe, NM (USA), `www.eyesopen.com/rocs`

Omega ein Konformationsensemble erstellt, dessen Mitglieder durch ROCS mit dem Referenzliganden überlagert werden. Die Überlagerung, die den besten Volumenüberlapp liefert, wird in der nächsten Phase dazu verwendet, den endgültigen Fitnesswert zu bestimmen. Er berechnet sich aus dem normalisierten Volumenüberlapp[66] zum Referenzliganden und der chemischen Ähnlichkeit, die über das Programm Triphic[10] bestimmt wird. Diese beiden Werte ergeben zusammen eine Zahl zwischen 0,0 und 2,0, die der EA-Inventor als Fitnesswert verwendet.

Der in dem EA-Inventor verwendete genetische Algorithmus ist ein regulärer populationsbasierter stochastische Algorithmus mit der Besonderheit, dass er eine einfach zugängliche Schnittstelle für benutzerdefinierte Bewertungsfunktionen wie in diesem Fall FLAP bietet. Die Grundlage für den chemischen Raum, den der Algorithmus durchsucht, liefert – wie auch bei Qsearch – ein chemischer Fragmentraum. Der Fragmentraum, der in der Publikation verwendet wurde, besteht aus über 1300 Fragmenten, die aus dem MDDR[11] Datensatz generiert wurden, und die über 32 Regeln neu verknüpft werden können.[12]

NovoFLAP wurde erfolgreich innerhalb von AstraZeneca im Rahmen einer Forschungskampagne für den $5HT_{1B}$- Rezeptor[13] verwendet, wobei neue Moleküle auf Basis eines fixen Startfragments entworfen wurden. Die Moleküle, die von dem Programm vorgeschlagen wurden, sind nachträglich von medizinischen Chemikern leicht modifiziert worden und dienten als Grundlage einer neuen Strukturklasse von nanomolaren Inhibitoren des $5HT_{1B}$- Rezeptors.[70]

3.4.3. BIBuilder

Das Programm BIBuilder[38] wurde 2011 von der Firma Böhringer Ingelheim vorgestellt. Es ist eine der wenigen Anwendungen, in denen das Problem des *de novo* Designs mit Berücksichtigung dreidimensionaler Informationen durch eine erschöpfende Suche gelöst wird. Abgesehen von FlexNovo[31] wird die erschöpfende Suche ansonsten hauptsächlich für die Rekombination von bekannten aktiven Molekülen (also in sehr kleinen Fragmenträumen) verwendet.[46] BIBuilder durchsucht den Fragmentraum im Rahmen einer Tie-

[10] Triphic; AstraZeneca Pharmaceuticals: Wilmington, DE (USA). Nicht veröffentlicht. Die Arbeitsweise von Triphic wird als analog zur CLIP[67] beschrieben, das Ähnlichkeitwerte auf Basis von Subgraphen bestimmt.

[11] MDL Drug Data Report; Symyx Technologies, Inc.: Sunnyvale, CA (USA). `www.accelrys.com`

[12] Der Datensatz wurde von Tripos und AstraZeneca entwickelt, aber nicht veröffentlicht.

[13] Der 5-Hydroxytryptamin 1B Rezeptor ist im zentralen Nervensystem lokalisiert. Seine Inhibition erhöht (indirekt) die Produktion von Osteoblasten, die in der Knochenproduktion involviert sind[68]. Eine erhöhte Aktivität hingegen wird mit Depressionen und dem Hang zum Drogenmissbrauch in Verbindung gebracht[69].

fensuche, wobei die generierten Moleküle aus maximal drei Fragmenten bestehen. Um mit der kombinatorischen Explosion umzugehen, werden zwei Strategien angewendet: die Reduktion der Berechnungskomplexität einerseits und eine frühe Einschränkung des Suchraums durch das Verwerfen von Teillösungen, die bestimmte Kriterien nicht erfüllen, andererseits.

Der Algorithmus zum Aufbau der Moleküle verfährt folgendermaßen: Jedes Eingabefragment wird in allen möglichen Orientierungen innerhalb des Pharmakophors als Startpunkt platziert. Diese Startplatzierung dient als Ausgangspunkt für das weitere Wachsen des Moleküls. An allen Anknüpfungspunkten dieses Fragments werden alle erlaubten Fragmente in verschiedenen Orientierungen angebaut, wobei ihre Konformationen jeweils mithilfe des Programms Omega berechnet werden. Nach jedem Anbauschritt werden die Teillösung bewertet und gegebenenfalls verworfen. Die Laufzeit der Gesamtmethode ist relativ hoch und bewegt sich zwischen 500 und 72,000 Computerstunden.

Die Komplexität der Berechnungen wird dadurch reduziert, dass sämtliche Teillösungen nur mit einem dreidimensionalen Pharmakophor bewertet werden. Im Vergleich zu aufwendigeren Scoring-Funktionen ermöglicht dies Geschwindigkeitsvorteile, da nur einfache geometrische Berechnungen notwendig sind, und die Anzahl der Kriterien für die Bewertung des Liganden niedrig gehalten wird. Um die Suche weiter zu beschleunigen, werden in jedem Anbauschritt Mindestanforderungen für die weitere Expansion einer Teillösung definiert. Erfüllt eine Teillösung diese nicht, so wird sie verworfen, wodurch sich frühzeitig Bereiche des Suchraums als nicht erfolgversprechend identifizieren und überspringen lassen.

Erst nachdem ein Molekül vollständig aufgebaut worden ist, wird es unter Berücksichtigung der gesamten Bindetasche mithilfe eines Kraftfeldes minimiert und anschließend durch eine Scoring-Funktion bewertet. Letztendlich werden sämtliche erzeugte Lösungen in einer Ergebnisliste zusammengefasst und dem Benutzer zurückgegeben.

Die Anwendbarkeit des Ansatzes wurde durch die Reproduktion von spezifischen Liganden aus großen Fragmenträumen gezeigt. So befand sich ein solcher Ligand für Renin unter 24.000 generierten Lösungen und für den Estrogenrezeptor konnte er unter 142.000 generierten Molekülen gefunden werden.

3.5. Neuheitswert der Methode

Die in Qsearch zum Einsatz kommende Kombination von dreidimensionalen Pharmakophoren mit dem stochastisch unterstützen Erstellen von Molekülen aus chemischen Fragmenträumen und einer dynamischen Ausrichtung ist in der Literatur bisher nicht be-

schrieben. Die ähnlichste Methode wurde im Programm SkelGen[30, 51] verwirklicht, die jedoch keine chemischen Fragmenträume verwendet, sondern einfache Kohlenstoffgerüste, die nachträglich dekoriert werden. Darüber hinaus besteht in dem SkelGen Algorithmus keine methodische Trennung zwischen der Suche im chemischen Raum und der Optimierung im Konformationsraum, wodurch der Suchraum beide Komponenten beinhaltet und dadurch stark vergrößert wird. Ein weiteres Alleinstellungsmerkmal von Qsearch ist die Entdekoration von aus Fragmenten gebauten Molekülen, die den Suchraum spezifisch um für ein Pharmakophor relevante Teile vergrößert.

4. Modellierung

In diesem Kapitel wird die Modellierung der eingesetzten chemischen Konzepte für die computergestützte Berechnung durch Qsearch erläutert. Dazu gehört das Modell, mit dem die Fragmenträume beschrieben werden, die Art und Weise in der die potentiellen Interaktionen von Molekülen bestimmt werden, die Beschreibung der Pharmakophore und im Rahmen eines kurzen Exkurses die Methode, die für die Bestimmung des molekularen Volumens zur Anwendung kommt.

4.1. Fragmentraummodell

Das in Qsearch verwendete Fragmentraummodell basiert auf den Vorarbeiten, die im Rahmen der Erstellung von FlexNovo[31] durchgeführt wurden, und in denen das in FTrees-FS[29] zum Einsatz kommende Modell um die Handhabung von dreidimensionalen Koordinaten erweitert wurde. Da Qsearch intern auf dem NAOMI-Atommodell[71] basiert, wurde das Fragmentraumkonzept aus FlexNovo leicht umformuliert und neu implementiert. Die Gemeinsamkeiten und Unterschiede der beiden Implementierungen werden im folgenden diskutiert.

Als Fragment wird ein Molekül bezeichnet, an dem an einer oder mehreren Stellen sogenannte Linkatome annotiert sind, die die Stellen markieren, an denen Verknüpfungen mit anderen Fragmenten durchgeführt werden können. Dabei beinhaltet jedes Linkatom selbst drei Informationen: Einen Linktyp, der beschreibt, mit welchen anderen Linkatomen Verknüpfungen durchgeführt werden dürfen, den *Exitvektor*[40], der die Position und Orientierung definiert, in der andere Fragmente angefügt werden können, und einem Bindungstyp, der die Valenz des zum Linkatom benachbarten Atoms definiert. Die Linktypen werden in der Regel analog zu den Schnittregeln, mit denen die Fragmente aus Molekülbibliotheken erstellt werden, vergeben, und der *Exitvektor* ergibt sich aus dem Winkel, in dem das Linkatom zu dem Fragment steht. Er wird genauso wie der Bindungstyp aus der geschnittenen Bindung übernommen.

Um das Konzept der Linkatome in das NAOMI-Modell von Element / Valenzzustand / Atomtyp zu integrieren, wurde ein neues Element („LINKER“) eingeführt, und drei neue

Valenzzustände („Lk_100", „Lk_010", „Lk_001") mit dazu korrespondierende Atomtypen („Lk_100", „Lk_010", „Lk_001") erstellt, die Informationen über die geschnittene Bindung beinhalten. Sie werden dazu verwendet, Linkatome, die beim Schneiden einer einfachen, doppelten oder dreifachen Bindung erstellt worden sind, zu kennzeichnen. So wird es ermöglicht, die Fragmente über die reguläre Molekülinitialisierung mithilfe der NAOMI-Routinen durchzuführen.

Ein Fragmentraum besteht aus einer Menge von Fragmenten und einem Satz an Regeln, die beschreiben, wie diese Fragmente verknüpft werden können. Die Unterschiede zu der Fragmentraumimplementierung aus FlexNovo ergeben sich größtenteils aus den implizit vorhandenen chemischen Informationen im NAOMI-Modell, was eine Vereinfachung der Fragmentraumdefinitionen zur Folge hat.[1] Folgende Punkte werden nicht mehr als Eingabe benötigt:

Torsionswinkel Bisher war es nötig, planare Geometrieren entlang von neu geknüpften Bindungen explizit kenntlich zu machen, damit sie korrekt erstellt werden. Das NAOMI-Modell hingegen erkennt Bindungen, die zu delokalisierten Systemen gehören, automatisch und konstruiert die Moleküle so, dass an diesen planare Winkel eingestellt werden. Dabei wird stets eine E-Konformation entlang der neu geknüpften Bindung ausgewählt.

Bindungslängen FlexNovo erlaubte die Bindungslänge für neu geknüpfte Bindungen als Eingabeparameter anzugeben. Das NAOMI-Chemiemodell stellt diese Informationen zur Verfügung. Sie wird daher nicht mehr als zusätzliche Eingabe benötigt.

Atomtyptransformationen Das FlexNovo-Molekülmodell verwendet für die Beschreibung kleiner Moleküle die Atomtypen, die im Tripos Kraftfeld[72] für Proteine definiert wurden. Dadurch, dass das Kraftfeld nicht die gesamte Spannbreite von Atomtypen in Wirkstoffmolekülen abdeckt, ist es teilweise nötig, die Zuordnung der Kraftfeldtypen zu den Atomen, die adjazent zu einer neu entstandenen Bindung liegen, zu verändern. Durch die differenziertere Abbildung der Atomtypen im NAOMI-Modell ist dieser Schritt nicht mehr nötig. Dies hat den zusätzlichen Vorteil, dass bestimmte Deskriptoren für Moleküle wie die Anzahl an Akzeptoren oder rotierbaren Bindungen nun über Fragmente hinweg additiv sind.

Amidbindungen Im Tripos Kraftfeld werden die delokalisierten Bindungen in Amiden separat behandelt. Für FlexNovo musste angegeben werden, dass Amidbindungen

[1] Um eine Abwärtskompatibilität der Fragmenträume zu gewährleisten, wird das Eingabeformat von FlexNovo weiterhin verwendet, allerdings werden die nun obsolet gewordenen Informationen ignoriert,

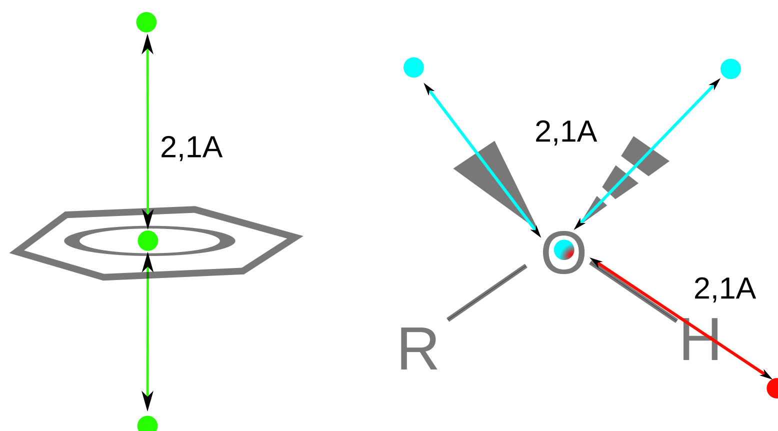


Abbildung 4.2.1.: Platzierung der Interaktionspunkte anhand von zwei Beispielen. Bei einem aromatischen Ring werden zwei Interaktionen generiert (grün) Beide beginnen im Ringmittelpunkt und zeigen senkrecht vom Zentrum weg. Bei Akzeptoren (blau) zeigen Interaktionen in Richtung der Elektronenpaare und bei Donoren zeigen die Interaktionen in Richtung der polaren Wasserstoffatome (rot).

bei einer Verknüpfung entstehen. Das NAOMI-Konzept kommt ohne diese explizite Modellierung aus.

4.2. Interaktionsmodell

In einem Pharmakophor sind die elektronischen Merkmale, die für die Ausbildung von supramolekularen Interaktionen notwendig sind, enthalten. Um ein Molekül mit den Anforderungen eines Pharmakophors zu vergleichen, müssen die potentiellen Interaktionen, die es ausbilden kann, beschrieben werden. Das in Qsearch implementierte Interaktionsmodell basiert auf den Arbeiten zu molekularen Interaktionen, die von Böhm für das Programm LUDI[7] durchgeführt worden sind.

In der Adaption von Qsearch werden ausschließlich gerichtete Interaktionen explizit modelliert. Hierbei besteht eine potentielle Interaktion immer aus zwei Punkten: einem Start- und einem Endpunkt. Die Distanz zwischen beiden Punkten ist für eine maximale Kompatibilität mit den MOE Pharmakophoren, die Qsearch als Eingabe dienen, auf 2,1Å festgelegt. Im folgenden wird auf die in Abschnitt 2.2 aufgezählten Interaktionstypen eingegangen, und deren genaue Zuweisung dargestellt.

(Wasserstoffbrücken-)Akzeptoren Als Akzeptor werden alle Stickstoff- und Sauerstoffatome betrachtet, die mindestens ein freies Elektronenpaar haben. Bestimmte Atome sind

explizit von dieser Zuweisung ausgenommen, da sie diese Interaktionen entweder gar nicht ausbilden, oder die ausgebildeten Interaktionen zu schwach sind, um im Pharmakophorkontext berücksichtigt zu werden. Dies sind alle Heteroatome von Nitrogruppen, die Stickstoffatome in Aminen, sowie die einfach gebundenen Sauerstoffatome in Ethern und Estern. Das Zentrum der Interaktion bildet jeweils das polare Schweratom, und die Richtungen werden durch die Position der freien Elektronenpaare vorgegeben.

(Wasserstoffbrücken-)Donoren Das NAOMI-Modell[71] stellt über seine Atomtypklassifikation die Information zur Verfügung, ob ein Atom als Donor fungieren kann oder nicht. Das Zentrum der Interaktion wird auf das Schweratom gesetzt, und die Richtungen werden jeweils durch die Position der an das Atom gebundenen Wasserstoffatome vorgegeben.

Aromatische Ringe Um die Interaktionen zu beschreiben, die durch einen aromatischen Ring ausgebildet werden können, werden drei Punkte verwendet, die zusammen zwei Interaktionsrichtungen beschreiben. Der Startpunkt beider Interaktionen wird in den geometrischen Mittelpunkt des Rings gelegt, und die zwei verschiedenen Endpunkte werden jeweils ober- und unterhalb der Ringebene platziert. Bei Ringsystemen wird dies für jeden einzelnen aromatischen Ring durchgeführt.

Weitere Interaktionstypen Zusätzlich zu den eben genannten gerichteten Interaktionstypen werden noch die Formalladungen und hydrophoben Regionen berücksichtigt, die aber nicht explizit modelliert werden. Sie werden anhand von Moleküleigenschaften in den jeweiligen Bewertungsfunktionen abgefragt. (vgl. Abschnitt 5.5)

4.3. Pharmakophormodell

Die Grundlage für die Bewertung von Molekülen in Qsearch stellt ein benutzerdefinierter Pharmakophor dar. In diesem Abschnitt werden die elektronischen Merkmale erläutert, die in Qsearch verwendet werden, um Pharmakophore zu beschreiben. Sie basieren auf dem zuvor erläuterten Interaktionsmodell und sind wiederum die Voraussetzung für die in Abschnitt 5.5 definierte Straffunktion, die in den stochastischen Optimierungsalgorithmen zum Einsatz kommt.

Pharmakophor	Zentrum	Richtung
Akzeptor	Schweratom	freies Elektronenpaar
Donor	Schweratom	gebundenes Wasserstoffatom
Aromatischer Ring	Mittelpunkt	Normale zur Ringebene
Positive Ladung	Atom mit Ladung	gebundenes Wasserstoffatom
Negative Ladung	Atom mit Ladung	freies Elektronenpaar
Hydrophob	unpolares Atom	

Tabelle 4.1.: Elektronische Merkmale von Pharmakophoren und ihre Entsprechungen in Molekülstrukturen.

4.3.1. Sterische Eigenschaften

Die sterische Anteil eines Pharmakophors definiert, welchen Raum ein Molekül innerhalb einer Zielstruktur einnehmen darf. Er wird in Qsearch über einen Satz Kugeln modelliert, was insbesondere Vorteile im ligandbasierten Design hat, wo die erlaubte Form anhand eines oder mehrerer überlagerter aktiver Moleküle abgeleitet wird. Dabei wird eine Kugel auf jedes Atom platziert, und so der Raum bestimmt, der für die Moleküle im aktiven Zentrum zur Verfügung steht.

In Qsearch erfüllt ein Molekül die sterischen Auflagen eines Pharmakophors dann, wenn keines seiner Schweratome außerhalb der Pharmakophorform liegt, was über die Atommittelpunkte bestimmt wird. Dass die Wasserstoffatome vernachlässigt werden, kann über ihre geringe Größe gerechtfertigt werden und ermöglicht bedeutende Geschwindigkeitsvorteile, die sich auch vergleichbare Methoden zu nutze machen.[18]

4.3.2. Elektronische Pharmakophormerkmale

Neben dem sterischen Anteil beinhaltet ein Pharmakophor auch elektronische Merkmale, die ebenfalls Bedingung für die *molekulare Erkennung* eines Moleküls sind. Diese Merkmale werden von Qsearch mithilfe von annotierten Kugeln modelliert. Sie beschreiben die räumliche Verteilung der elektronischen Eigenschaften, die ein Molekül aufweisen muss, um dem Pharmakophor zu entsprechen. Die elektronischen Eigenschaften entstammen einem Konsens, der zwischen den etablierten Methoden gebildet hat, und zuerst in der Beschreibung des Programms Catalyst[73] formuliert wurde.[74] (vgl. Abschnitt 2.3)

Die Pharmakophormerkmale sind eng mit dem Interaktionsmodell verbunden, wobei sich für die meisten Merkmale eine 1:1 Beziehung zu den Interaktionstypen herstellen lässt. Die Annotationen der Kugeln beziehen sich auf die potentiellen Interaktionen, die ein Molekül ausbilden kann, wobei eine Kugel sich entweder auf den Start- oder

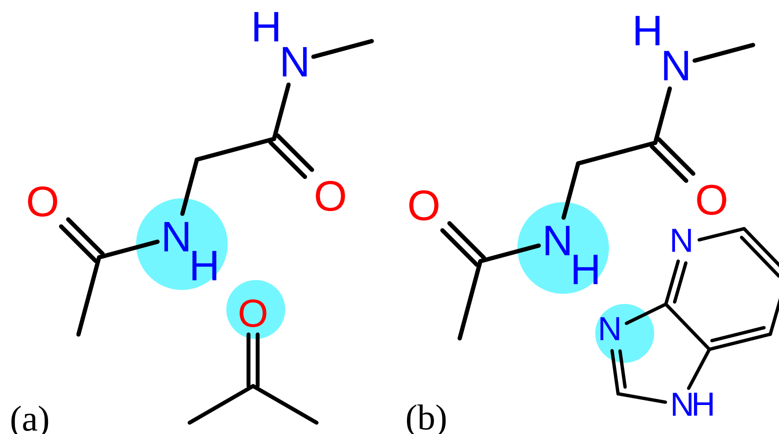


Abbildung 4.3.1.: Ein Pharmakophor für die Interaktion mit einem Proteinrückgrad (cyan) und zwei verschiedene molekulare Fragmente die ihn erfüllen. (a) ist ein erwünschtes Fragment, da die Richtung der Interaktion eine Wasserstoffbrückenbindung erlauben würde, (b) ist ein eher unerwartetes Fragment, da die beiden Pharmakophormerkmale von verschiedenen Atomen erfüllt werden.

Endpunkt einer Interaktion bezieht. Ein elektronisches Pharmakophormerkmal ist dann erfüllt, wenn ein entsprechender Punkt in ihm liegt. Eine Zusammenstellung der modellierten Eigenschaften und ihren korrespondierenden molekularen Motiven ist in Tabelle 4.1 aufgeführt.

Einen Spezialfall im implementierten Pharmakophormodell stellt die Modellierung der hydrophoben Eigenschaften dar. In Qsearch werden Atome als hydrophob angesehen, wenn sie keine Formalladung tragen und weder Wasserstoffbrückendonor noch -akzeptor sind. Ein hydrophobes Pharmakophormerkmal ist dann erfüllt, wenn mindestens ein hydrophobes Atom in ihm liegt, und dort gleichzeitig kein polares Atom vorhanden ist. Eine ausführliche Diskussion der Handhabung von hydrophoben Pharmakophormerkmalen findet sich in Abschnitt 6.5.1.

4.3.3. Verknüpfung von Pharmakophormerkmalen

Um die Richtung einer gewünschten Interaktion mithilfe eines Pharmakophors zu definieren, reicht es nicht aus, zwei voneinander unabhängige Merkmale zu verwenden. So könnten zwei Pharmakophormerkmale, die einen Start- und Endpunkt definieren, jeweils von verschiedenen Molekülteilen erfüllt werden, wenn dies nicht explizit vorgegeben ist. (vgl. Abbildung 4.3.1) Um diesen Fall zu vermeiden, können Merkmale verknüpft werden, wodurch gefordert wird, dass sie von Interaktionen erfüllt werden müssen, die den

gleichen Startpunkt teilen.

Prinzipiell lassen sich nur Pharmakophormerkmale vom gleichen Typ verknüpfen. Eine Ausnahme bilden die Merkmale, die Ladungen definieren. Negative Ladungen können mit Akzeptor-Richtungen, und positive Ladungen mit Donor-Richtungen kombiniert werden. Dies spiegelt die Definition eines geladenen Amins oder einer geladenen Carbonsäure mitsamt ihren Ausrichtungen wider.

Die Anzahl der Merkmale, die verknüpft werden können, ist nicht auf zwei beschränkt. Ein häufiges Beispiel ist die Verknüpfung eines Akzeptor-Zentrums mit zwei Akzeptor-Richtungen, um die Position einer Carbonylgruppe zu beschreiben. In diesem Fall können nur Atome, die mindestens zwei freie Elektronenpaare besitzen, dieses kombinierte Merkmal erfüllen.

4.4. Molekulare Flexibilität und Konformationen

Moleküle haben in der Regel keine starren Form, sondern zeichnen sich durch geometrische Flexibilität aus. An sogenannten „rotierbaren Bindungen“ können sich Molekülteile gegeneinander verdrehen, wodurch eine Menge an möglichen dreidimensionalen Formen, den sogenannten Konformationen, für ein Molekül entsteht. Dabei gibt es Faktoren, die die freie Rotierbarkeit an Bindungen einschränken. In der Regel ist dies die sterische Hinderungen durch große Molekülteile, die bei einer Rotation kollidieren würden. Aber auch andere Faktoren können für eine Erhöhung der Energiebarriere verantwortlich sein. Hierzu zählen delokalisierte π-Elektronensysteme zwischen den Atomen die an der Bindung beteiligt sind, und die Fixierung einer Bindung durch Ringschlüsse.[2]

In Qsearch kommt ein vereinfachendes Modell für die molekulare Flexibilität zum Einsatz. In dem Modell wird jede Bindung als rotierbar betrachtet, die folgende Bedinungen erfüllt:

1. sie ist eine Einzelbindung
2. sie ist keine Ringbindung
3. sie ist nicht Teil eines delokalisierten Elektronensystems
4. sie ist nicht terminal
5. sie verbindet keinen Methyl- oder Trifluormethylrest

[2]Darüber hinaus gibt es an bestimmten molekularen Mustern Winkel, die präferiert angenommen werden. Eine weiterführende Diskussion findet sich in Abschnitt 6.5.2.

6. sie verbindet kein Nitril

In diesen Kriterien kommen zusätzlich zu den chemischen Voraussetzung für die Rotierbarkeit auch Heuristiken für häufige symmetrische Gruppen zusammen, durch die der zu durchsuchende Konformationsraum eines Moleküls ohne den Verlust von relevanter Information verkleinert werden kann.

4.5. Molekulares Volumen

Das van-der-Waals Volumen eines Moleküls ist das Volumen, das von ihm eingenommen wird, und in dem keine Atome anderer Moleküle Platz finden.[75] Es spielt in der Wirkstoffforschung – gerade bei QSAR Ansätzen – eine entscheidende Rolle. Zum Beispiel ist es unter anderem einer der Faktoren, der die Geschwindigkeit des passiven Transports in die Zellen bestimmt.[76] Im Kontext von Qsearch ist allerdings noch wichtiger, dass Moleküle, die ein Volumen haben, das zu groß für die sterischen Beschränkungen eines Pharmakophors ist, diesen immer verletzen, und daher ohne Weiteres übersprungen werden können. (vgl. Schnelle Ablehnung in Abschnitt 5.2)

In der Regel wird das Volumen eines Moleküls über die Vereinigung von atomzentrierten Kugeln beschrieben, deren Radius dem van-der-Waals Radius des jeweiligen Elements entspricht.[77] Das Problem bei dieser Berechnungsmethode liegt für Qsearch darin, dass die Koordinaten der Atome vorliegen müssen, um sie anwenden zu können. Die Bestimmung der Koordinaten ist jedoch aufwendig, weswegen es wünschenswert ist, eine Methode zu verwenden, die ohne dreidimensionale Informationen auskommt.

Aus verschiedenen Publikationen geht hervor, dass sich das Volumen von Molekülen sehr gut auf Basis der Molekültopologie vorhersagen lässt. Beispielsweise können Beiträge von Molekülfragmenten summiert[75], oder auch einfache Faustformeln[78] dafür verwendet werden, das van-der-Waals Volumen anzunähern. Vor diesem Hintergrund wurde für die Anwendung in Qsearch eine Parametrisierung der NAOMI-Atomtypen[71] in Bezug auf ihre Volumenbeiträge vorgenommen. Dazu wurde der ZINC „lead-like“ Datensatz[79] in der Version 8 mit 2,7 Millionen Molekülen und annotierten dreidimensionalen Koordinaten verwendet. Die einzelnen Moleküleinträge liegen in verschiedenen Protonierungszuständen vor, was eine umfassendere Berücksichtigung von verschiedenen Atomtypen ermöglicht, als bei der Verwendung von Molekülen mit Standardprotonierungen, bei denen zum Beispiel alle Carbonsäuren deprotoniert sind.

Ein Nachteil dieses Datensatzes ist, dass die Koordinaten mithilfe eines Kraftfelds automatisch generiert wurden. Dies führt zu einer Parametrisierung auf das zum Einsatz gekommene Chemiemodell (in diesem Fall das der Firma OpenEye) und nicht auf die phy-

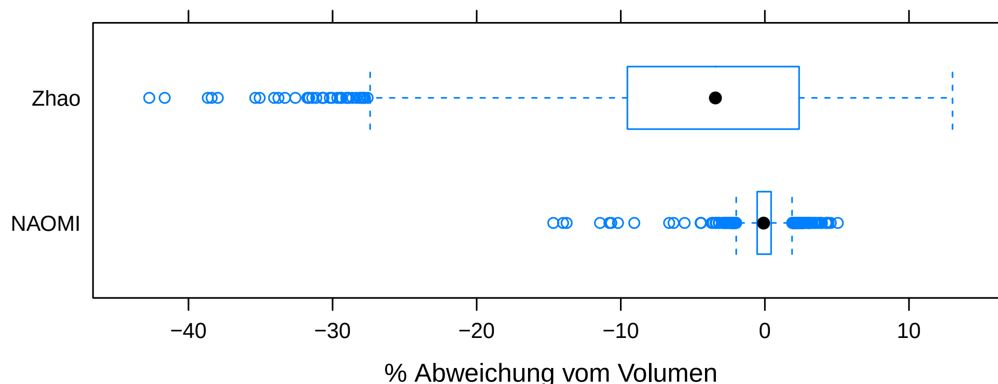


Abbildung 4.5.1.: Vergleich der Abweichung vom nach Grant und Pickup[77] geometrisch ermittelten Volumen nach der Methode von Zhao und nach den NAOMI-Atomtypen in einem Box-Whisker Plot. Die verwendeten Moleküle entstammen dem WDI Datensatz aus 2008.

sikalisch korrekten Werte. Allerdings sollte die Größenordnung der Unterschiede zwischen verschiedenen Chemiemodellen bei der Verwendung in Qsearch nicht ins Gewicht fallen. Diese Annahme wird durch den Vergleich mit von CORINA[80] generierten Koordinaten in Abbildung 4.5.2 bekräftigt.

Entsprechend der Hypothese, dass die Atombeiträge unabhängig voneinander zum Volumen beitragen, wurde eine lineare Regressionsanalyse auf den Daten angewendet. Dafür wurden aus den 2,7 Millionen Molekülen zufällig 100.000 als Trainingsdatensatz ausgewählt. Für jedes Molekül wurde das Volumen geometrisch über die Vereinigung der Atomkugeln berechnet, und die jeweilige Atomtypverteilung bestimmt. Diese Daten wurden als Grundlage für eine Analyse der Atomtypenbeiträge mittels QR-Zerlegung verwendet. Das entstandene Modell zeigte einen r^2-Wert von 0,9987, was für einen sehr guten Zusammenhang zwischen den Atomtypen und den Volumen steht. Die einzelnen Atomtypenbeiträge sind in Anhang C aufgeführt.

Um das Verhalten der Methode im Vergleich zu anderen beurteilen zu können, wurde das Modell von Zhao[78] als Alternative herangezogen.[3] Das Modell berechnet das Volumen für ein Molekül M auf Basis der Volumenbeiträgen der einzelnen Atomkugeln und kompensiert ihre Überschneidungen mithilfe der Anzahl der Bindungen (B), sowie der Anzahl aromatischen (R_{A}) und nicht-aromatischen (R_{NA}) Ringsysteme, $vdw(a)$ liefert das van-der-Waals Volumen eines Atoms:

[3] Für die Auswahl ausschlaggebend war die Verfügbarkeit (bzw. Nichtverfügbarkeit) von alternativen Methoden und Datensätzen, die nur kommerziell verfügbar sind.

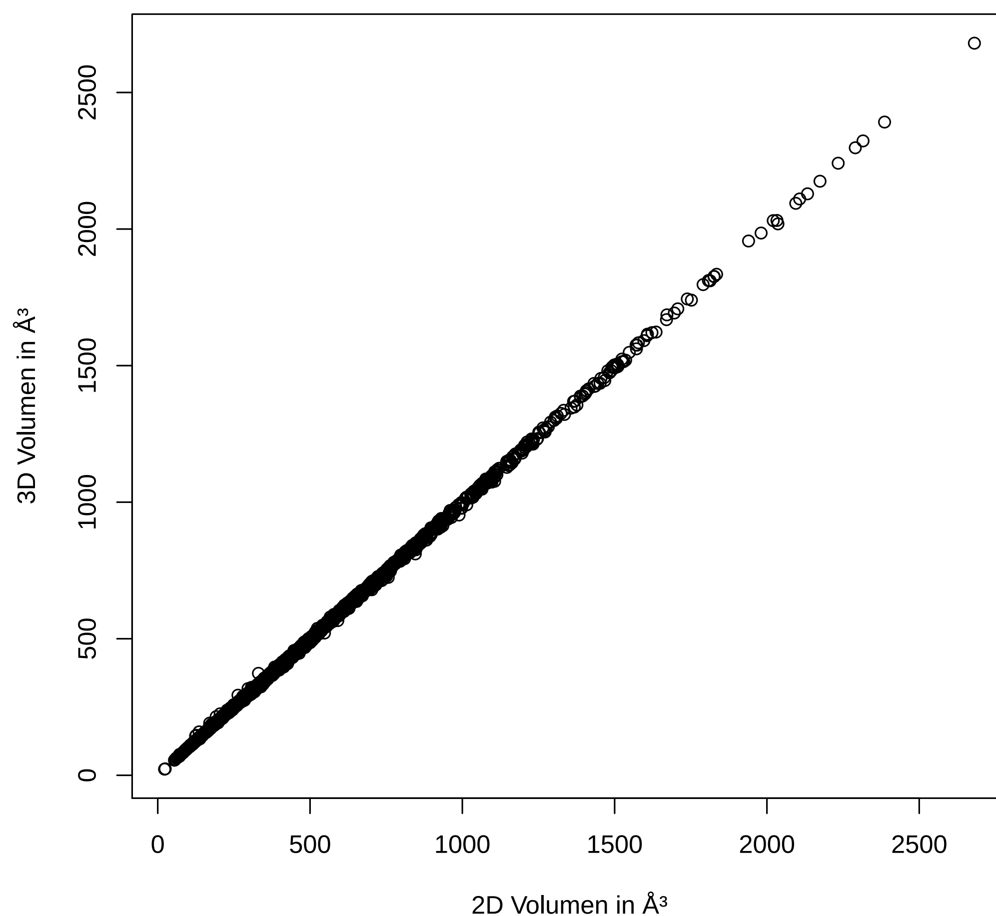


Abbildung 4.5.2.: Vergleich zwischen dem Volumen, das mithilfe der Atomtypen berechnet wurde, und dem Volumen, das geometrisch bestimmt wurde.

$$V_{\mathrm{vdW}}\left(M\right) = \sum_{a \in M} \frac{4\pi}{3} vdw(a)^3 - 5,92B - 17,4R_{\mathrm{A}} - 3,8R_{\mathrm{NA}}$$

Als Testdatensatz wurde der WDI Datensatz[4] aus 2008 verwendet, der 4295 Moleküle enthält, für die mit dem Programm CORINA[80] 3D-Koordinaten erstellt wurden. Auf dieser Basis wurde das Volumen geometrisch[77], mittels der NAOMI-Atomtypen, und der Formel nach Zhao berechnet. In Abbildung 4.5.2 sind die Abweichungen der jeweiligen Methode von der geometrischen Berechnung aufgetragen, die als Vergleichswert verwendet wurde. Es ist ersichtlich, dass die Vorhersage auf Basis der NAOMI-Atomtypen weitaus weniger Abweichung von den geometrisch ermittelten Werten mit sich trägt und darüber hinaus eine geringere Varianz bei den Abweichungen aufweist. Folglich wurde für die Volumenberechnung in Qsearch dieser Ansatz ausgewählt.

[4]World Drug Index; Thomson: Philadelphia, PA (USA).

5. Methoden

In diesem Kapitel werden die Methoden beschrieben, die im Rahmen des Projektes zum Einsatz gekommen sind. Dafür wird zuerst ein allgemeiner Überblick über den Arbeitsablauf in Qsearch gegeben und daraufhin auf spezifische Teilaspekte der Gesamtmethode eingegangen.

5.1. Allgemeiner Überblick über die Methode

In der Literatur für das *de novo* Design existieren verschiedene Ansätze, um Fragmenträume nach bestimmten Kriterien zu durchsuchen. Für die Methoden, die lediglich auf zweidimensionalen Molekülbeschreibungen basieren, wurden neben stochastischen Verfahren (vgl. [60, 81, 82]) auch Ansätze veröffentlicht, die die Komplexität des Problems algorithmisch angehen und es optimal lösen (vgl. FTrees-FS[29]). Vergleichbare algorithmische Lösungen für dreidimensionale Molekülbeschreibungen, wie sie in Qsearch verwendet werden, finden sich in der Literatur hingegen nicht[1], da es bisher nicht gelungen ist, das Problem so zu formulieren, dass es ist in vertretbarer Zeit optimal gelöst werden kann.

Die Algorithmen, die in Qsearch zum Einsatz kommen, verwenden stochastische Methoden, wobei dem Simulated Annealing[83] eine herausragende Bedeutung zukommt. Diese Methode eignet sich für große Suchräume mit unstetigen Energielandschaften, und wurde oft bei vergleichbaren Problemen verwendet.[48]

Der Qsearch Algorithmus beginnt mit der zufälligen Generierung von Molekülen aus den Eingabefragmenten, bis eines gefunden worden ist, das von der Größe und von den Interaktionen her den Pharmakophor erfüllen kann. Ausgehend von diesem Startpunkt tritt Qsearch dann in eine Schleife von kontinuierlichen molekularen Modifikationen ein. In jedem Schritt werden die Fragmente des aktuellen Moleküls modifiziert, und das daraus resultierende Molekül an dem Pharmakophor ausgerichtet. Abhängig davon, wie gut

[1] Ausnahmen bilden die Programme FlexNovo[31] und BIBuilder[38]. Sie lösen das Problem durch eine Aufzählung aller möglichen Lösungen. Trotz ihrer Heuristiken zur Beschränkung des Suchraums zeichnen sie sich durch lange Laufzeiten und großen Ressourcenbedarf aus. Qsearch hingegen wurde explizit für begrenzte Rechner-Ressourcen geschrieben, damit es auf üblichen Desktop Computern ausgeführt werden kann.

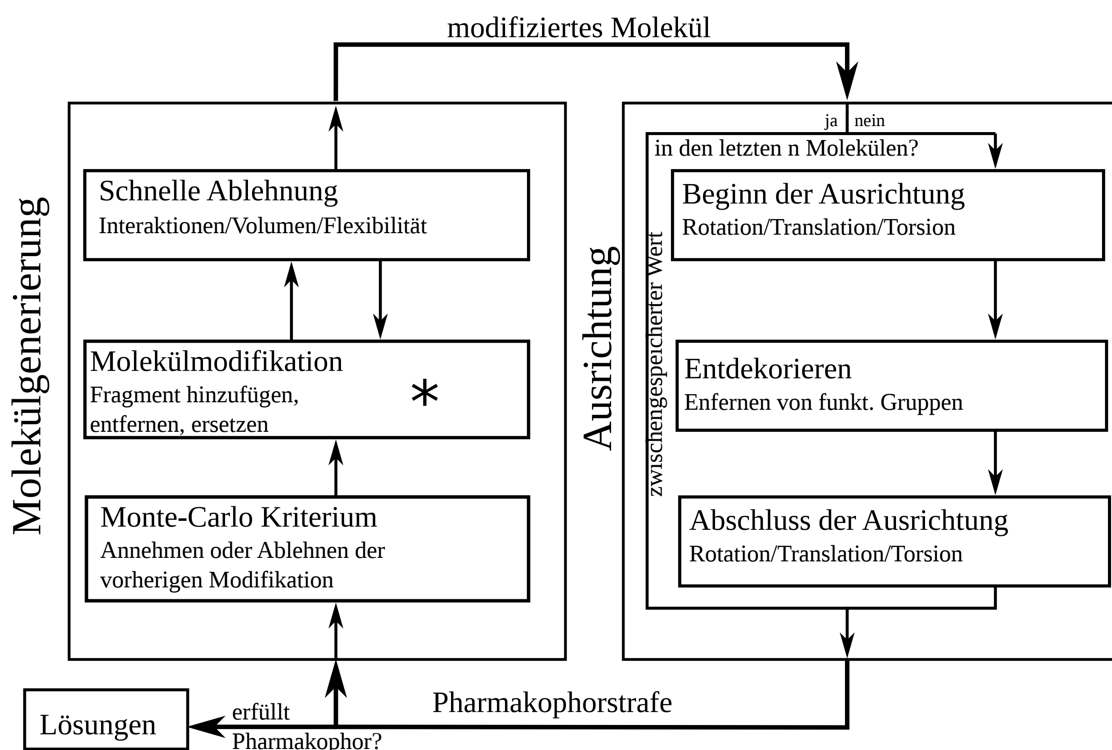


Abbildung 5.1.1.: Allgemeiner Überblick über den Ablauf des in Qsearch verwendeten Algorithmus. Der Algorithmus beginnt mit einem zufällig generierten Molekül an der mit einem * markierten Stelle und vollführt eine definierte Anzahl von vollständigen Schleifendurchläufen. Zu Beginn der Ausrichtungsphase wird überprüft, ob das Molekül unter den letzten n ausgerichteten Molekülen war, und diese gegebenenfalls übersprungen. Wenn das ausgerichtete Molekül den Pharmakophor erfüllt, wird es als Lösung ausgegeben.

der Pharmakophor erfüllt werden kann, wird die Modifikation entweder akzeptiert und als Grundlage für den nächsten Schritt verwendet, oder abgelehnt, wobei der nächste Schleifendurchlauf dann mit dem Molekül aus der letzten Iteration erneut durchgeführt wird. Wenn ein Molekül gefunden wird, das alle Pharmakophorbedingungen erfüllt, so wird es als Lösung ausgegeben. Der Algorithmus bricht ab, sobald eine vorgegebene Anzahl an Schleifendurchläufen ausgeführt worden ist.

5.2. Modifikation von Molekülen und schnelle Ablehnung

In dem von Qsearch verwendeten Algorithmus werden kontinuierlich neue Moleküle aus Teillösungen erzeugt. Dafür wird ein Satz an Operationen verwendet, mit denen Modifikationen an Molekülen durch die Veränderung ihrer zugrunde liegenden Fragmente durchgeführt werden.[2] Die Operation, die in jedem Schritt konkret durchgeführt wird, wird jeweils zufällig bestimmt, wobei die relative Häufigkeit der einzelnen Operationen frei eingestellt werden kann.[3] Ein Molekül kann durch folgende Operatoren verändert werden:

Hinzufügen An einem freien Linkatom des Moleküls wird ein Fragment hinzugefügt.

Entfernen Ein terminales Fragment des Moleküls wird entfernt.

Ersetzen Ein beliebiges Fragment des Moleküls wird durch ein anderes ersetzt.

Nicht jedes Molekül ist für alle Pharmakophore interessant und/oder wirkstoffähnlich. Daher wurde ein Satz an schnellen Ablehnungskriterien erstellt, um Modifikationen, die in solchen Molekülen resultieren, schnell zurückzuweisen. So kann die aufwendige Ausrichtung an dem Pharmakophor übersprungen werden, was einen positiven Einfluss auf die Geschwindigkeit des Programms hat. Die Kriterien, die für die schnelle Ablehnung Anwendung finden, sind hier zusammengefasst und werden sowohl in Bezug auf ihre Relevanz und ihre Berechnung erläutert.

Volumen Die Größe der Moleküle wird über das Volumen in benutzerdefinierten Grenzen gehalten. Es wird wie in Abschnitt 4.5 beschrieben über die NAOMI-Atomtypen berechnet, so dass keine Koordinaten für die Berechnung notwendig sind. Der Aufwand

[2] Zusätzliche Modifikationen auf der Ebene von funktionellen Gruppen werden im Entdekorierungsschritt durchgeführt. (vgl. Abschnitt 5.4)

[3] Dabei hat es sich als praktikabel erwiesen, die relativen Häufigkeiten des Hinzufügens und des Entfernens von Fragmenten gleich zu halten, da ansonsten die Moleküle stetig anwachsen beziehungsweise schrumpfen.

der Ausrichtung von Molekülen an den Pharmakophor wächst quadratisch mit der Größe der Moleküle. Daher ist es wünschenswert, Moleküle zu überspringen, die ohnehin zu groß für das geforderte Volumen sind. Zu kleine Moleküle werden vom Benutzer in der Regel als nicht relevant betrachtet, da mithilfe des Pharmakophorvolumens implizit auch die Größe der erwarteten Lösungen beschrieben wird, und sollten ebenfalls übersprungen werden. Die Definition des erlaubten Rahmens für das Volumen wird durch den Benutzer vorgenommen und relativ zum Gesamtvolumen des Pharmakophors angegeben. Als Empfehlung für die Benutzer werden Werte von 80%-120% des Eingabepharmakophors verwendet.[4]

Flexibilität Um die Flexibilität der erhaltenen Moleküle in Bereichen zu halten, die bei der Grundgerüstsuche sinnvoll sind, lässt sich die Länge des maximalen Pfades aufeinander folgender rotierbarer Bindungen im Molekülgraphen beschränken. Moleküle, deren Pfad zu lang ist, werden sofort abgelehnt. Dies geschieht aus zwei Gründen. Erstens hat die Flexibilität über die Entropiebeiträge einen Einfluss auf die Affinität von Wirkstoffen, das heißt, Moleküle, die sehr flexibel sind, haben eine niedrigere Bindungsaffinität als ähnliche starre Moleküle.[84] Zweitens wächst auch die Komplexität der Ausrichtung an den Pharmakophor mit der Flexibilität der Moleküle, da jede frei drehbare Bindung im Molekül den Suchraum vergrößert. In der Regel werden maximal zwei aufeinander folgende rotierbare Bindungen erlaubt.

Interaktionen Moleküle, die nicht genügend Interaktionen ausbilden können, um alle Pharmakophormerkmale zu erfüllen, werden ebenfalls sofort abgelehnt. In den mit Qsearch verwendeten BRICS-Fragmenträumen tritt dieser Fall allerdings nicht sehr häufig auf, da die in ihnen enthaltenen Fragmente stark dekoriert sind. Es gibt jedoch auch molekulare Muster, die relativ selten in den Fragmenten vorkommen. Dazu gehören sp^2-hybridisierte Akzeptoratome, und vor allen Dingen Gruppen mit Formalladungen. Ein schnelles Ablehnen von Molekülen ohne diese Funktionalitäten ermöglicht es, den Suchraum auf für den jeweiligen Pharmakophor relevantere Teile zu beschränken.

[4] Dass das Volumen höher ist als das des Pharmakophors ist der Handhabung der sterischen Eigenschaften geschuldet. Da dabei nur die Atommittelpunkte überprüft werden, und nicht die Kugeln, die die Ausdehnung der Atome beschreiben, können Moleküle, deren Atome an der Grenze der erlaubten Form liegen, ein größeres Volumen als der Pharmakophor selbst haben.

5.3. Ausrichtung an den Pharmakophor

Um Aussagen darüber machen zu können, ob ein Molekül einen Pharmakophor erfüllen kann, muss sein Konformationsraum mit Hinblick auf die Kompatibilität zu dem Pharmakophor überprüft werden. Hierfür gibt es eine Reihe an computergestützten Verfahren, deren Methoden sich in zwei verschiedene Ansätze aufteilen lassen. Für den ersten Ansatz wird eine möglichst repräsentative Menge von Konformationen erzeugt, die starr mit dem Pharmakophor überlagert werden. Diese Methode kommt zum Beispiel in LigandScout[85] und NovoFLAP[65] zum Einsatz. Im zweiten Ansatz wird der Konformationsraum der Moleküle dynamisch im Rahmen einer Optimierung durchsucht und die dreidimensionale Orientierung sowie die Torsionswinkel verändert, um zu einer möglichst guten Überlagerung zu gelangen. Diese Optimierung kann entweder numerisch wie beim „Directed Tweak"[86], oder auch stochastisch wie im Programm SkelGen[30] durchgeführt werden.

Die beiden Ansätze zeichnen sich durch unterschiedliche Stärken aus.[74] Der erstere erlaubt das Aufteilen des Problems in einen Vorverarbeitungsschritt (die Generierung der Konformationen) und in einem Suchschritt (die konkrete Suche), wodurch die Daten vorverarbeitet werden können, so dass ein effizienter Zugriff ermöglicht wird. Darüber hinaus lässt sich so die nicht-triviale Aufgabe der Konformationsgenerierung an eine spezialisierte externe Methode delegieren. In diesem Ansatz ist eine repräsentative Generierung der Konformationsensembles von großer Wichtigkeit. Fehlende relevante Konformationen führen zu einer falsch negativen Bewertung von Molekülen da ihre relevanten Konformationen nicht enthalten sind.

In zweiten Ansatz werden keine Vorberechnungen angestellt, was in einer langsameren Laufzeit Ausdruck findet. Allerdings wird es durch die flexiblere Auswahl der Torsionswinkel auch ermöglicht, ungewöhnliche Konformationen zu finden, die zwar energetisch nicht optimal, aber trotzdem valide sind.[87]

In Qsearch wird der zweite Ansatz verwendet. Jedes Molekül, das generiert wird und den schnellen Ablehnungskritieren genügt, wird daraufhin untersucht, wie gut es den Pharmakophor erfüllen kann. Dazu muss der Raum der möglichen Konformationen des Moleküls durchsucht werden, was in Qsearch mithilfe eines stochastischen Algorithmus durchgeführt wird. Ein Vorverarbeitungsschritt bietet sich in diesem Szenario nicht an, da jedes Molekül in der Regel nur ein einziges Mal betrachtet wird.

Bevor ein Molekül an einem Pharmakophor ausgerichtet wird, wird überprüft, ob es unter den letzten n ausgerichteten Molekülen war.[5] Ist dies der Fall, so wird das zwi-

[5] n ist ein frei wählbarer Parameter. In den hier präsentierten Berechnungen wurden n so gewählt, dass

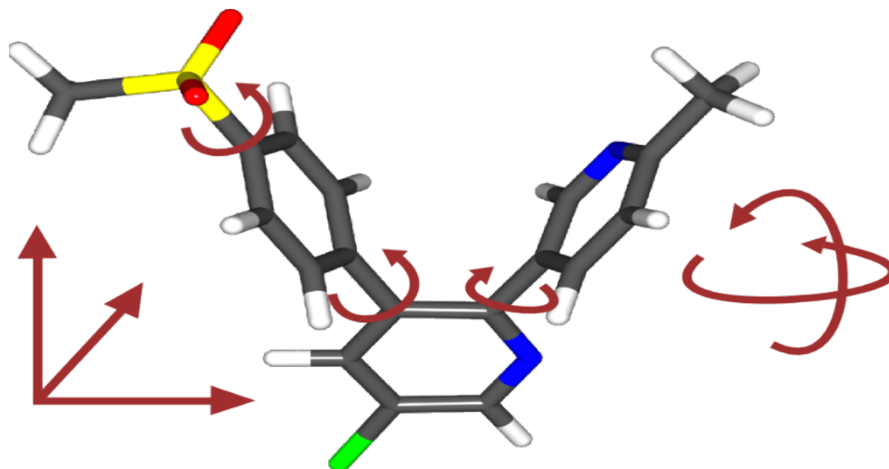


Abbildung 5.3.1.: Freiheitsgrade bei der Optimierung der Konformation (in rot): Translation und Rotation im Raum sowie Veränderung der Torsionswinkel an rotierbaren Bindungen.

schengespeicherte Ergebnis der letzten Ausrichtung wiederverwendet. Es beinhaltet die numerische Bewertung des Moleküls, das Molekül, das durch den Entdekorierungsschritt entsteht, und die jeweils dazugehörigen Koordinaten. Die Identität der Moleküle wird über den USMILES[88] bestimmt, der über die NAOMI-Implementierung[71] mit den zusätzlich annotierten Stereoinformation berechnet wird.

5.3.1. Stochastische Optimierung der Konformation

Bei der Suche nach einer möglichst optimalen Konformation wird wie auch bei der Molekülgenerierung auf die Methode Simulated Annealing[83] zurückgegriffen. Der Algorithmus beginnt mit einer Konformation, die möglichst nahe an der des Vorgängermoleküls liegt, was dadurch erreicht wird, dass die Koordinaten der unveränderten Fragmente wenn möglich erhalten bleiben und nur für die Atome der neuen Fragmente berechnet werden. In der darauf folgenden iterativen Optimierung werden die Koordinaten des Moleküls durch die folgenden Operationen verändert:

Translation Das Molekül wird um einen Vektor verschoben, wobei die Richtung des Vektors über einen zufälligen Punkt auf der Einheitskugel bestimmt wird. Die Länge des Vektors wird zufällig so gewählt, dass sie kleiner oder gleich einem benutzerdefinierten Maximalwert ist.

Rotation Das Molekül wird um eine Achse gedreht, die durch den geometrischen Mole-

alle Moleküle eines Simulated Annealing Laufs zwischengespeichert werden.

külmittelpunkt und einen zufälligen Punkt auf einer Einheitskugel um ihn bestimmt wird. Der Winkel wird ebenfalls zufällig aus dem Bereich $[-x, +x]$ gewählt, wobei x ein benutzerdefinierter Maximalwert ist.

Torsionsänderung Der Torsionswinkel an einer zufälligen rotierbaren Bindung wird um einen zufälligen Betrag verstellt. Als rotierbar gelten in Qsearch alle Bindungen, die den in Abschnitt 4.4 aufgeführten Bedingungen genügen. Auch hier liegt der Winkel in dem Bereich $[-x, +x]$.

Welche dieser Operationen ausgewählt wird, ist zufällig, wobei die relative Häufigkeit der drei Möglichkeiten frei einstellbar ist. Nach einer definierten Anzahl an Schritten wird die beste Konformation, die während des Suchvorgangs generiert wurde, zurückgegeben.

Da die Methode stochastisch ist, kann nicht garantiert werden, dass der gefundene Wert global gesehen optimal oder nahe am Optimum ist. Um die Wahrscheinlichkeit zu reduzieren, dass eine qualitativ schlechte Lösung erhalten wird, wird der Algorithmus mit jedem Molekül mehrmals durchgeführt und nur der beste Wert aus allen Läufen verwendet. In den bisherigen Anwendungen hat sich ein Wert von vier Läufen bewährt. Die einzelnen Optimierungsläufe können unabhängig voneinander durchgeführt werden und lassen sich daher gut parallelisieren, was auf modernen Rechnersystemen große Geschwindigkeitsvorteile bringt.

5.3.2. Parametrisierung der stochastischen Optimierung

Um eine gute Ausrichtung der generierten Moleküle an den Pharmakophor zu gewährleisten, wurde eine Studie zur Parametrisierung der stochastischen Optimierung durchgeführt. Dafür wurde ein Datensatz von Molekülen[89] verwendet, die aus der „Protein Data Bank“[90] (PDB) extrahiert und über ihre Bindungstaschen zusammengefasst worden sind. Für jeden dieser Komplexe wurde – sofern möglich und sinnvoll – ein Pharmakophor erstellt und die dazugehörigen Moleküle mit verschiedenen Parametersätzen ausgerichtet. Die Parametersätze wurden dann auf Basis der Anzahl der korrekt ausgerichteten Molekülen bewertet.

Da die Anzahl der möglichen Parameterkombinationen sehr groß ist, und vor der Studie bereits viel Erfahrungswissen vorhanden war, wurde ihre Auswahl und Bewertung manuell ausgeführt. Als Ausgangspunkt wurden Parameter auf Basis der existierenden Erfahrungswerte gewählt.

Datensatz Der verwendete Datensatz wurde von Chan und Labute[89] für die Parametrisierung einer Bewertungsfunktion für die Überlagerung von kleinen Molekülen zusam-

mengestellt. Der Datensatz besteht aus Liganden, die aus der PDB nach einer Reihe von Kriterien ausgewählt worden sind:

- weniger als 300 Schweratome
- weniger als zehn Aminosäuren, wenn es sich um ein Peptid handelt
- Auflösung des gesamten Protein-Ligand Komplexes von mindestens 2,5Å
- nur organische Elemente
- kein Häm oder Zucker
- kein ATP-Analog
- mindestens ein Ringschluss im Molekül (um Fette zu entfernen)
- mindestens 2 Schweratome, die durch drei Bindungen getrennt sind (um Wasser und sehr kleine Moleküle zu entfernen)

Die verbliebenen Moleküle wurden über die Proteindomänen in denen sie gebunden vorliegen zusammengefasst.[6] In einem letzten Schritt wurden dann alle Liganden mit mehr als zwölf rotierbaren Bindungen verworfen, und sich zu ähnliche Moleküle aus den jeweiligen Datensätzen entfernt. Insgesamt ergab sich dabei eine Menge von 100 überlagerten Molekülgruppen.

Für die Parametrisierung in Qsearch wurde für jede der Molekülgruppen des Datensatzes mithilfe des Programms MOE[7] manuell ein Pharmakophor erstellt, was ungefähr bei der Hälfte der Molekülgruppen möglich war. Bei den restlichen gab es entweder nicht genügend „starke“ Interaktionen, wie sie für Pharmakophore üblich sind, war das Verhältnis von Volumen zu elektronischen Eigenschaften zu groß, oder die Überlagerung wurde durch ATP-analoge Gruppen oder Zucker dominiert.[8] Sämtliche Moleküle des Datensatzes, die nicht auf ihre zugehörigen Pharmakophore passen, wurden verworfen.

Auswahl der Parameter In jedem Schritt wurden vier verschiedene Einstellungen variiert, von denen jeweils zwei Einfluss auf den Simulated Annealing Algorithmus hatten und zwei auf die durchgeführten Konformationsänderungen. Mit jedem dieser Parametersätze

[6] Die Details der Zusammenfassungsprozedur sind in [89] ausführlich beschrieben, würden an dieser Stelle jedoch den Rahmen sprengen.

[7] MOE; Chemical Computing Group: Montreal (Kanada). `www.chemcomp.com`

[8] Die Autoren des Datensatzes haben anscheinend nur reine Zucker und sehr nahe ATP-Analoga ausgeschlossen. Glykolisierte Moleküle oder welche, den für ATP charakteristischen Purinring tragen, wurden im Filterprozess nicht berücksichtigt.

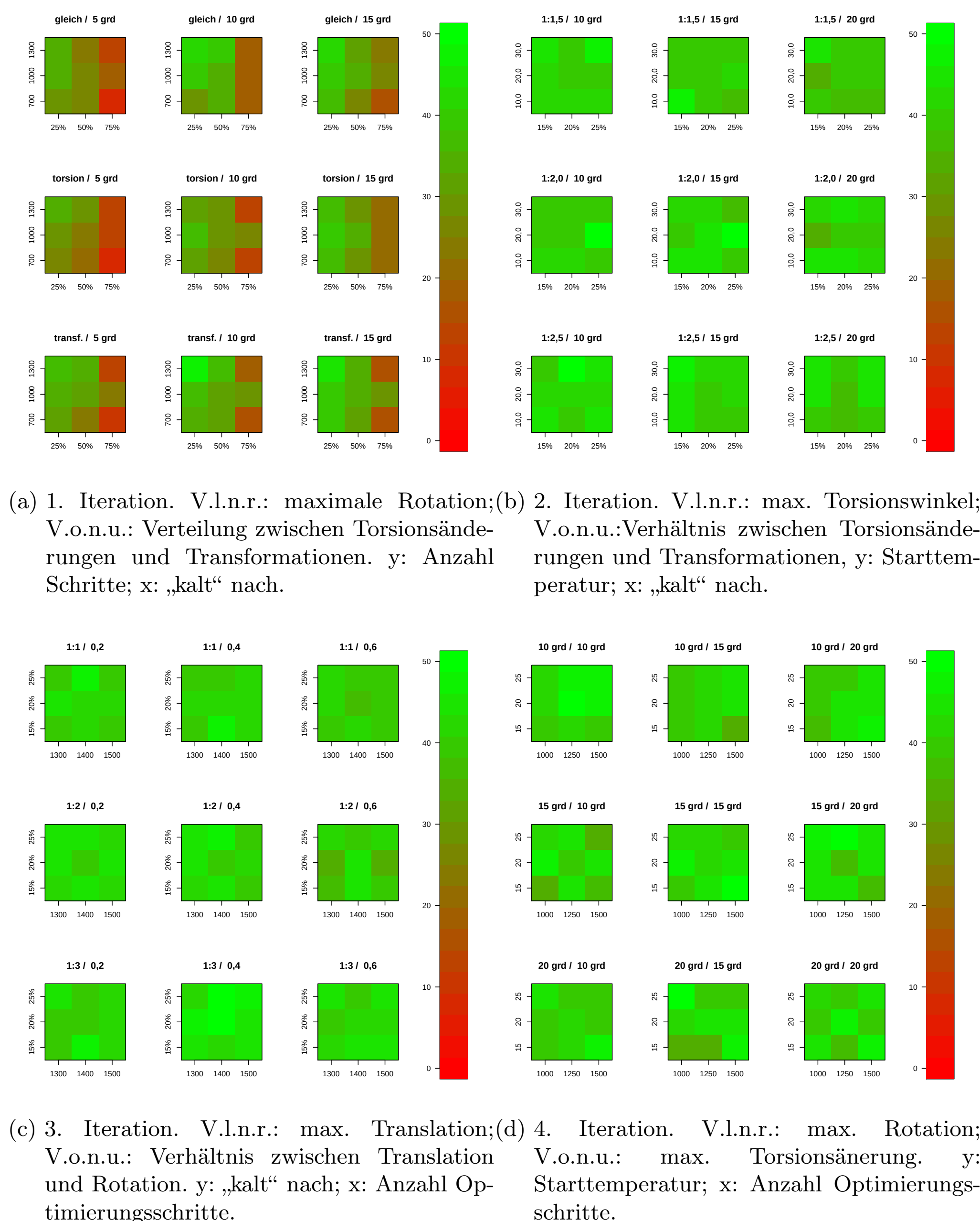
wurden sämtliche Moleküle des Datensatzes an ihrem jeweiligen Pharmakophor ausgerichtet und die Anzahl der korrekt ausgerichteten Moleküle farblich in einem Diagramm markiert. (vgl. Abbildung 5.3.2) Auf Basis einer visuellen Auswertung dieses Diagramms wurden dann die Parameter für den nächsten Lauf ausgewählt. Die Parameter für jeden einzelnen Lauf finden sich in Anhang D.

Erste Iteration Im ersten Schritt wurden verschiedene Abkühlungsstrategien angewendet, wobei die Abkühlungsfaktoren α jeweils so ausgewählt wurden, dass die Temperatur in der Optimierung nach 25%, 50% oder 75% der Zeit auf einen Wert nahe Null sinkt und nur noch Schritte akzeptiert werden, die zu einer Verbesserung der Bewertung geführt haben. (In den Abbildungen auch als „kalt“ bezeichnet.) Außerdem wurden verschiedene Laufzeiten für die Optimierung gewählt (700, 1000 und 1300 Schritte). Bei den Konformationsänderungen wurden verschiedene Verhältnisse zwischen Änderungen der Torsionswinkel und der Transformation der gesamten Konformation verwendet, wobei das Verhältnis von Rotation und Translation bei 1:3 fixiert war. (mehr Torsionen, gleiches Verhältnis, mehr Transformationen)

In dem Diagramm lässt sich erkennen, dass die Ergebnisse sich mit kürzeren Abkühlungsstrategien verbessern, weswegen für die weiteren Schritte α so gewählt wurde, dass die .Optimierung nach 25% der Laufzeit einen „kalten“ Zustand erreicht Eine längere Laufzeit wirkte sich in den meisten Fällen ebenfalls positiv aus, weswegen 1300 Schritte für das weitere Vorgehen verwendet wurden. Genauso ist zu sehen, dass mehr Transformationen als Torsionsänderungen besser waren als eine Gleichverteilung oder ein umgekehrtes Verhältnis, und dass ein maximaler Rotationswinkel von 10° zu guten Ergebnissen führt.

Zweite Iteration In diesem Schritt wurden verschiedene Starttemperaturen eingestellt, schnellere Abkühlungszeitpläne verwendet, und die maximalen Torsionsänderungen variiert. Da sich in der vorherigen Iteration zeigte, dass mehr Transformationen als Torsionsänderungen erfolgreicher sind, wurden darüber hinaus verschiedene Verhältnisse zwischen Transformationen und Torsionsänderungen verwendet. (1:1,5, 1:2,0, 1:2,5)

In dem Diagramm lässt sich erkennen, dass ein Verhältnis von Transformationen zu Torsionsänderungen von 1:2,5 kombiniert mit einer maximalen Torsionsänderung von 10° die meisten positive Ergebnisse liefert (Teildiagramm links unten), und sich eine Starttemperatur von 20,0 über alle Parametervariationen am besten auswirkt. Eine eindeutige Aussage zu dem Abkühlungsverlauf ließ nicht machen, weswegen er in den nachfolgenden Iterationen weiterhin untersucht wurde.



(a) 1. Iteration. V.l.n.r.: maximale Rotation; V.o.n.u.: Verteilung zwischen Torsionsänderungen und Transformationen. y: Anzahl Schritte; x: „kalt“ nach.

(b) 2. Iteration. V.l.n.r.: max. Torsionswinkel; V.o.n.u.:Verhältnis zwischen Torsionsänderungen und Transformationen, y: Starttemperatur; x: „kalt“ nach.

(c) 3. Iteration. V.l.n.r.: max. Translation; V.o.n.u.: Verhältnis zwischen Translation und Rotation. y: „kalt“ nach; x: Anzahl Optimierungsschritte.

(d) 4. Iteration. V.l.n.r.: max. Rotation; V.o.n.u.: max. Torsionsänerung. y: Starttemperatur; x: Anzahl Optimierungsschritte.

Abbildung 5.3.2.: Die Ergebnisse aus der Parametrisierung der Ausrichtung von Molekülen an den Pharmakophor. Insgesamt wurden die Parameter in vier Iterationen verbessert, wobei jeweils vier verschiedene Parameter variiert wurden. Zwei Parameter wurden für die Einteilung in neun Einzelgraphen verwendet und die weiteren für die Aufteilung der Einzelgraphen eingesetzt. Für jede Parameterkombination wurde die Zahl der korrekt ausgerichteten Moleküle farbkodiert in dem Graphen aufgetragen.

Dritte Iteration Für die dritte Iteration wurden erneut verschiedene Abkühlungsstrategien („kalt“ nach 15%, 20%, 25%) erprobt, sowie die Anzahl an Schritten zwischen 1300, 1400 und 1500 variiert. Bei den Konformationsänderungen wurden unterschiedliche Verhältnisse zwischen Rotation und Translation erkundet sowie die maximale Translation aus 0,2Å, 0,4Å und 0,6Å ausgewählt.

Insgesamt waren die meisten korrekt ausgerichteten Moleküle im Diagramm in der 3. Zeile und 2. Spalte zu finden, was darauf schließen lässt dass eine maximale Translation von 0,4Å kombiniert mit einem Verhältnis von Rotation zu Translation von 1:3 gute Ergebnisse liefert. Allerdings ließen sich erneut keine Aussagen den Abkühlungsstrategien und der Anzahl an Optimierungsschritten machen.

Vierte Iteration In der letzten Iteration wurden die Maximalwerte für die Torsionsänderungen und Rotationen zwischen 10°, 15° und 20° variiert und die Schrittzahl (1000, 1250, 1500) sowie die Starttemperatur (15, 20, 25) verändert.

In den Ergebnissen lässt sich ablesen, dass 10° sowohl für die maximale Torsionsänderung als auch die maximale Rotation zu den meisten korrekt ausgerichteten Molekülen führt (Teildiagramm links oben), und sich ein Wert von 20 bei der Starttemperatur als vorteilhaft erweist. Da die Ergebnisse von 1250 und 1500 Schritten vergleichbar sind, wurde eine Optimierungsdauer von 1250 Schritten festgelegt, um den Rechenaufwand in Qsearch zu beschränken.

Finale Parameter Für die finalen Parameter wurden bei den Konformationsänderungen ein Verhältnis für Translation:Rotation:Torsionsänderungen von 1:3:1,6 und maximale Änderungen von 0,4Å, 10° und 10° respektive gewählt. Für die stochastische Optimierung haben sich eine Starttemperatur von 20, eine Schrittzahl von 1250 und ein α-Wert von 0,988 bewährt.

5.4. Nachträgliche Entdekoration von Molekülen

Qsearch zeichnet sich durch einen gesonderten Entdekorierungsschritt während der Ausrichtungsroutine aus. In diesem Schritt werden funktionelle Gruppen, die bei einem Molekül nicht für die Erfüllung des Pharmakophors vonnöten sind, ebenso wie überflüssige Heteroatome aus aromatischen Ringen entfernt. Dadurch wird die Menge an Molekülen, die in einem Fragmentraum enthalten ist, dynamisch um solche Moleküle erweitert, die für einen spezifischen Pharmakophor besonders interessant sind.

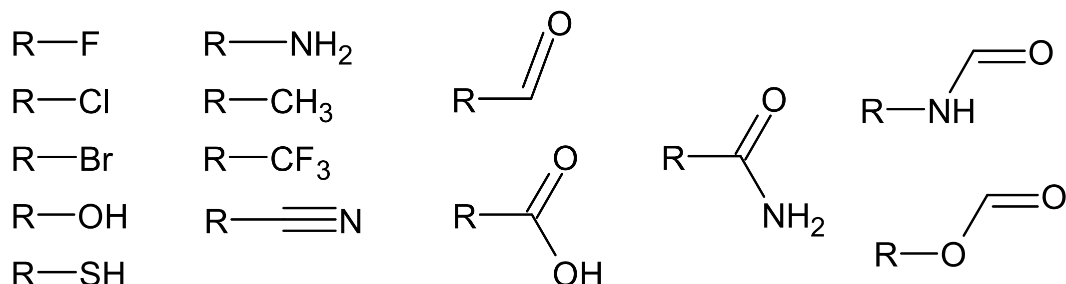


Abbildung 5.4.1.: Funktionelle Gruppen, die von Qsearch automatisch identifiziert und entfernt werden, wenn sie nicht für den Pharmakophor benötigt werden.

Dieser Schritt hilft, mit Fragmenträumen umzugehen, die aus dem Zerlegen von Bibliotheken bekannter Wirkstoffe erstellt wurden. Dieser Ursprung zeigt sich in den einzelnen Fragmenten unter anderem dadurch, dass sie stark mit polaren funktionellen Gruppen dekoriert sind, die in die Wirkstoffe eingeführt wurden, um ihre Löslichkeit zu verbessern. Darüber hinaus ist die Anzahl von Heteroatomen in aromatischen Heterozyklen in den verwendeten Fragmenträumen erhöht, was auf die Fragmentauswahlprozedur zurückzuführen ist. Die Überrepräsentierung von Heterozyklen in den BRICS-Räumen wird zusammen mit der Polarität in Abschnitt 6.1 diskutiert.

Sowohl die hohe Anzahl an funktionellen Gruppen als auch die starke Verbreitung der Heterozyklen führen dazu, dass die aus den Fragmenten resultierenden Moleküle sehr polar sind, und von medizinischen Chemikern als Benutzern des Programms sofort als nicht geeignet zurückgewiesen werden. Um die Anwenderakzeptanz zu steigern, wurden zwei Strategien entwickelt, um den oben genannten Problemen entgegenzuwirken und die im folgenden erläutert werden: Das Entfernen von funktionellen Gruppen und die Verringerung der Anzahl von Heteroatomen in aromatischen Ringen.

5.4.1. Identifikation und Eliminierung entbehrlicher funktioneller Gruppen

Der erste Teil des Verfahrens um Molekülen zu entdekorieren, ist die Eliminierung entbehrlicher terminaler funktioneller Gruppen. Dieser Schritt dient mehreren Zwecken: Zum einen können Moleküle, bei denen funktionelle Gruppen entfernt wurden, die sterischen Vorgaben des Pharmakophors besser erfüllen, da sie kleiner sind. So ist denkbar, dass es für ein Molekül erst möglich wird, einen Pharmakophor zu erfüllen, nachdem aus der erlaubten Form „herausstehende“ funktionelle Gruppen entfernt worden sind. Zum anderen verbessert der Schritt die Akzeptanz der Lösungen bei den Anwendern, was einerseits durch das Entfernen von funktionellen Gruppen, die nicht für den Pharmakophor not-

wendig sind, erreicht wird, und andererseits durch das Entfernen von ungewöhnlichen Resten, die durch das Verknüpfen von Fragmenten entstehen.

Das Entfernen der funktionellen Gruppen wird in einem iterativen Verfahren vorgenommen. Jede terminale funktionelle Gruppe wird daraufhin überprüft ob sie für die Erfüllung eines Pharmakophormerkmals benötigt wird, und andernfalls aus dem Molekül entfernt. Dieser Schritt wird für jedes Molekül so lange wiederholt, bis keine funktionelle Gruppe mehr entfernt werden kann. Dadurch werden auch zusammengesetzte Gruppen wie zum Beispiel lange Kohlenstoffketten oder Methoxygruppen durch den Algorithmus abgedeckt. Eine Auflistung aller funktionellen Gruppen, die durch den Algorithmus entfernt werden, ist in Abbildung 5.4.1 zu finden.[9]

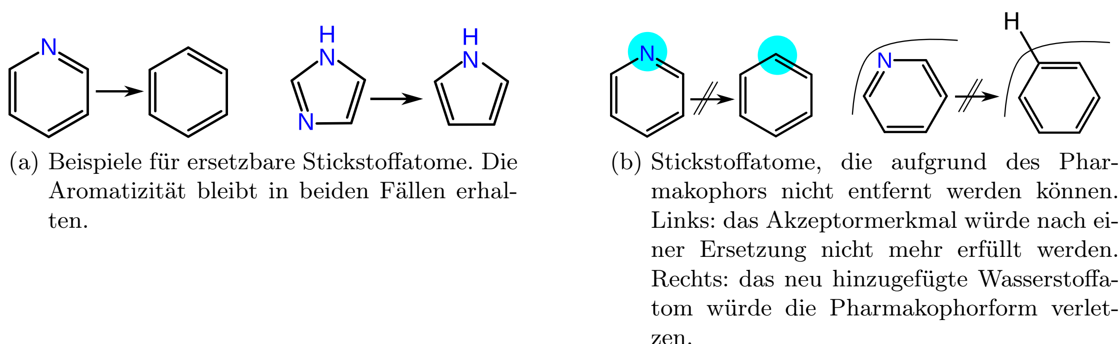
Voraussetzung um festzustellen, welche Gruppen für die Erfüllung eines Pharmakophors wichtig sind, und welche Gruppen im Entdekorierungsprozess entfernt werden können, ist eine grobe Ausrichtung des Moleküls an den Pharmakophor. In Qsearch findet daher die Entdekorierung nach 50% der Ausrichtung statt, um mit hoher Wahrscheinlichkeit auf einer Konformation zu arbeiten, in der Molekül das bereits so orientiert ist, dass erkennbar ist, welche funktionellen Gruppen zu welchen Pharmakophormerkmalen korrespondieren.

5.4.2. Verringerung der Anzahl von Heteroatomen in Heterozyklen

Ein weiterer Teil der Entdekorierungsstrategie, die in Qsearch zum Einsatz kommt, ist die Verringerung der Anzahl von Heteroatomen in aromatischen Heterozyklen. Dabei werden Heteroatome in aromatischen Ringen durch Kohlenstoffatome ersetzt, wenn der Pharmakophor dies zulässt und keine sterischen Probleme durch die Wasserstoffatome entstehen, die für die Absättigung der neu eingeführten Kohlenstoffatome angefügt werden müssen.

Die Heteroatome, die ersetzt werden sollen, müssen dabei sowohl in Bezug auf ihre Geometrie als auch auf ihre Valenzen zu Kohlenstoff äquivalent sein. Ersteres trifft für alle Heteroatome zu, da in regulären aromatischen Systemen alle beteiligten Schweratome eine planare Geometrie besitzen. Bei der Valenz sind allerdings weitergehende Einschränkungen zu beachten. Kohlenstoff hat in der Kekuléformeln-Repräsentation von aromatischen Ringsystemen zwei Einfachbindungen und eine Doppelbindung. Betrachtet man die Heteroatome, die in aromatischen Heterozyklen vorkommen, kann nur Stickstoff eine Einfach- und eine Doppelbindung aufweisen. Sauerstoff zum Beispiel kann hingegen nur

[9]Funktionelle Gruppen, die als Substruktur in anderen funktionellen Gruppen enthalten sind, werden nicht entfernt. Daher kommen in der Abbildung einige Motive öfter und/oder als Teile von anderen vor. Die Aufteilung in funktionellen Gruppen wird durch das NAOMI-Molekülmodell[71] zur Verfügung gestellt.



(a) Beispiele für ersetzbare Stickstoffatome. Die Aromatizität bleibt in beiden Fällen erhalten.

(b) Stickstoffatome, die aufgrund des Pharmakophors nicht entfernt werden können. Links: das Akzeptormerkmal würde nach einer Ersetzung nicht mehr erfüllt werden. Rechts: das neu hinzugefügte Wasserstoffatom würde die Pharmakophorform verletzen.

Abbildung 5.4.2.: Entfernen von Stickstoffatomen aus aromatischen Heterozyklen.

zwei Einfachbindungen ausbilden.

Die Atome, die für eine Ersetzung geeignet sind, lassen sich leicht über ihre zugewiesenen Valenzzustände identifizieren, da in dem NAOMI-Molekülmodell den aromatischen Systemen stets Kekuléformeln zugrunde liegen. Alle Stickstoffatome in aromatischen Systemen mit dem Valenzzustand N(110)-Aro können durch C(210)-Aro Kohlenstoffatome ersetzt werden. Die zweite Einfachbindung des Kohlenstoffs wird mit einem neu hinzugefügten Wasserstoffatom abgesättigt. Der Atomersetzungsschritt findet direkt nach der Eliminierung unnötiger funktionellen Gruppen statt, um die Wahrscheinlichkeit zu verringern, dass die die hinzugefügten Wasserstoffatome mit anderen Atomen kollidieren.

5.5. Bewertung von Konformationen

Sowohl für die Suche im Fragmentraum als auch für die Optimierung von Konformationen ist es nötig, Moleküle und Konformationen in Bezug auf die Erfüllung eines Pharmakophors zu vergleichen. Qsearch verwendet eine numerische Funktion $p_{\text{score}}(M, P)$ um ein Molekül M in Bezug auf einen Pharmakophor P zu bewerten. Diese Funktion berücksichtigt zwei Aspekte in ihrer Beurteilung. Zum einen wird die geometrische Validität der Konformation bewertet, und zum anderen die Erfüllung des Pharmakophors durch das Molekül.

$$p_{\text{score}}(M, P) = p_{\text{conf}}(M) + p_{\text{pharm}}(M, P)$$

Bewertung der internen Energie einer Konformation Ziel des Terms $p_{\text{conf}}(M)$ ist es, physikalisch invalide Konformationen von Molekülen zu vermeiden, wobei sich die Validität von Konformationen von Molekülen in Bezug auf mehrere Aspekte bewerten lässt.

Für miteinander verbundene Atome werden dazu klassischerweise die Bindungslänge, der Bindungs- und der Torsionswinkel herangezogen, und für nicht miteinander verbundene Atome die sterische Abstoßung und die elektrostatische Anziehung.[91]

Qsearch beschränkt sich bei der Bewertung der Konformationen nur auf eine Kollisionsüberprüfung der Atome. (Die Einschränkungen, die dadurch entstehen, werden unter anderem in Abschnitt 6.5.2 diskutiert.) Für den Bewertungsterm von Konformationen ergibt sich folglich:

$$p_{\text{conf}}(M) = p_{\text{clash}}(M)$$

In der Regel wird der Kollisionsterm in Molekulardynamiksimulationen durch das Lennard-Jones Potential modelliert, das sowohl die Anziehung als auch die Abstoßung von Atomen auf Basis ihrer Distanz d beschreibt, wobei ε und σ variable Parameter sind:

$$\nu(d) = 4\varepsilon \left[\left(\frac{\sigma}{d}\right)^{12} - \left(\frac{\sigma}{d}\right)^{6} \right]$$

Die Einführung eines Anziehungsterms in Qsearch führt dazu, dass die Moleküle im Laufe der Optimierung „kollabieren“ und kompakte Formen ergeben. Da in Qsearch die Molekülform ohnehin durch den Pharmakophor vorgegeben wird, kann der Anziehungsterm in der Bewertung vernachlässigt werden, und es dem Benutzer über die Pharmakophordefinition überlassen werden, sich für kompakte oder ausgedehnte Konformationen zu entscheiden. Der Abstoßungsterm auf der anderen Seite entbehrt in seiner konkreten Formulierung einer „starken theoretischen Grundlage“[91] und wird daher nur in einer vereinfachten Form verwendet, die sich am Lennard-Jones Potential orientiert, aber schneller zu berechnen ist. Wenn die Distanz zweier Atome a und b zu klein ist, wird sie folgendermaßen bestraft:

$$p_{\text{clash}}(a,b) = \begin{cases} \left(\frac{vdw(a)+vdw(b)}{d(a,b)}\right)^2 & \text{wenn } d(a,b) < vdw(a) + vdw(b) \\ 0.0 & \text{sonst} \end{cases}$$

$vdw(a)$ ergibt den van-der-Waals Radius eines Atoms, und $d(a,b)$ die Distanz zwischen den Mittelpunkten zweier Atome. Die gesamte Strafe für ein Molekül wird über alle Paare von Atomen gebildet. Ausgenommen sind Atome, die nur durch ein oder zwei Bindungen getrennt sind, oder die sich im gleichen Ringsystemen befinden, beziehungsweise im Falle von Wasserstoffatomen, die an das gleiche Ringsystem gebunden sind.[10]

[10] Da Qsearch Ringkonformationen nicht verändert, wird bei allen Ringsystemen der Eingabe davon ausgegangen, dass sie eine valide Konformation besitzen und ihre interne Energie nicht neu berechnet.

$$p_{\text{clash}}(M) = \sum_{a,b \in M, a \neq b} p_{\text{clash}}(a, b)$$

Bewertung der Pharmakophorerfüllung Der zweite Teil der Straffunktion von Qsearch ist $p_{\text{pharm}}(M, P)$ und beschreibt die geometrische Abweichung des Moleküls zum Pharmakophor. Analog zur Pharmakophordefinition setzt sich dieser Term aus einem sterischen und aus einem elektronischen Teil zusammen. Zusätzlich existiert noch ein Term für die Bewertung der hydrophoben Merkmale.

$$p_{\text{pharm}}(M, P) = p_{\text{st}}(M, st(P)) + p_{\text{el}}(M, el(P)) + p_{\text{hyd}}(M, hyd(P))$$

$st(P)$ liefert den sterischen Anteil des Pharmakophors, $el(P)$ die elektronischen Merkmale und $hyd(P)$ die hydrophoben Merkmale.

Die Strafe für die Verletzung der Pharmakophorform wird über alle Atome summiert. Dabei wird für jedes Atom über $d(a, s)$ die Distanz zur nächsten Kugel berechnet, mit der die Form beschrieben worden ist. Wenn ein Atom in einer Kugel liegt, ist die Distanz als 0 definiert:

$$p_{\text{st}}(M, S) = \sum_{a \in M} \min \left\{ d(a, s)^2 | s \in S \right\}$$

Die Strafe für die Nichterfüllung von elektronischen Pharmakophormerkmalen wird in der Funktion $p_{\text{el}}(M, el(P))$ berechnet. Die Gesamtstrafe wird über alle einzelnen Merkmale aufsummiert, wobei jeweils nur der Wert zu der Interaktion mit der minimalen Strafe betrachtet wird.

$$p_{\text{el}}(M, el(P)) = \sum_{e \in E} \min \left\{ \left(\frac{d(i, e)}{r(e)} \right)^2 | i \in ia(M, e) \,, \infty \right\}$$

$ia(M, e)$ liefert die Interaktionen eines Moleküls M, die zu einem Pharmakophormerkmal e kompatibel sind, $d(i, e)$ liefert die Distanz eines Interaktionspunktes i zu dem Mittelpunkt des Pharmakophormerkmals e und $r(e)$ seinen Radius.

Die Strafe für die hydrophoben Merkmale setzt sich aus zwei Komponenten zusammen. Zum einen wird mit $p_{\text{hyd/hyd}}(M, H)$ bestraft, wenn sich kein hydrophobes Atom innerhalb des Merkmals befindet, und mit $p_{\text{hyd/pol}}(M, H)$ alle hydrophilen Atome bestraft, die innerhalb des Merkmals liegen:

$$p_{\text{hyd}}(M, H) = p_{\text{hyd/hyd}}(M, H) + p_{\text{hyd/pol}}(M, H)$$

$p_{\text{hyd/hyd}}(M, H)$ errechnet sich analog zu $p_{\text{el}}(M, E)$, nur dass statt der Koordinaten der Interaktionen die der hydrophoben Atome verwendet werden:

$$p_{\text{hyd/hyd}}(M, H) = \sum_{h \in H} \min \left\{ \left(\frac{d(a, h)}{r(h)} \right)^2 | a \in h(M) \, , \infty \right\}$$

$h(M)$ ergibt die hydrophoben Atome eines Moleküls, $d(h, a)$ die Distanz eines Atoms zu dem Mittelpunkt eines hydrophoben Merkmals und $r(h)$ seinen Radius.

Die Strafen für Atome, die in hydrophoben Merkmalen liegen ergibt sich wie folgt:

$$p_{\text{hyd/pol}}(M, H) = \sum_{h \in H} \sum_{a \in M \setminus h(M)} \left\{ f_2 \left(\frac{d(h, a)}{r(h)} \right) \right\}$$

Hierbei werden alle Atome, die nicht hydrophob sind, bestraft, wenn sie in hydrophoben Merkmalen liegen. In der Funktion wird die sigmoide Straffunktion f_2 verwendet, um die (normierte) Nähe eines Punktes zum Kugelmittelpunkt zu bestrafen.

$$f_2(d) = \frac{1}{1 + e^{10 \cdot (d - 0{,}75)}}$$

5.6. Suche nach kompatiblen Fragmenten beim Ersetzungsschritt

Die Ersetzung eines bestehenden Fragments durch ein anderes, das äquivalent in Bezug auf seine Linkatome ist, ist im Molekülmodifikationsschritt eine der Operationen, mit denen ein Molekül verändert werden kann. Um diese Operation auszuführen, ist es notwendig, alle Fragmente aus dem Fragmentraum zu bestimmen, die diese Äquivalenz bieten. Dabei treten drei algorithmische Fragestellung auf, die in den unten stehenden Ausführungen erläutert werden.

1. Das Auffinden der Kombination von Linktypen, die ein Fragment aufweisen muss, um äquivalent zu einem anderen zu sein.

2. Das effiziente Suchen von Fragmenten mit einem bestimmten Satz an Linktypen.

3. Das Erzeugen der möglichen Verknüpfungen eines Fragments mit dem Restmolekül.

Das Fragment, das ersetzt werden soll, wird im folgenden als f bezeichnet. Der Molekülgraph, das durch das Entfernen von f entsteht, wird als *Restmolekül* bezeichnet, und die durch diese Entfernung neu geöffneten Linkatome des Restmoleküls werden durch das Symbol L repräsentiert.

5.6.1. Bestimmung der kompatiblen Linktypkombinationen

Der erste Schritt ist die Bestimmung der Kombinationen von Linktypen, die es einem Fragment erlauben, f zu ersetzen. Ein Fragment g kann f dann ersetzen, wenn es mindestens die Linkatome besitzt, um mit allen Linkatomen des Restmoleküls L entsprechend den Regeln des Fragmentraums verknüpft werden zu können. Dabei muss beachtet werden, dass ein Linkatom des Restmoleküls zu mehreren anderen Linktypen kompatibel sein kann.

Die Erzeugung aller möglichen Kombinationen aus einem Satz von möglichen Elementen ist ein bekanntes Problem, dessen Lösung in Algorithmus M („mixed-radix generation“) von Knuth[92] beschrieben wird. Der folgende Algorithmus 5.1 basiert auf dieser Lösung und adaptiert sie für das vorliegende Problem. Der Algorithmus generiert das kartesische Produkt aller Mengen von Linkatomen, die jeweils zu einem Atom in L kompatibel sind.

5.6.2. Verwendung eines Bitmapindexes für Fragmentraumanfragen

Nachdem mit dem oben aufgeführten Algorithmus 5.1 die Kombinationen von Linktypen bestimmt wurden, die ein Fragment besitzen muss, um f ersetzen zu können, müssen alle Fragmente aus dem Fragmentraum gefunden werden, die eine dieser Kombinationen von Linkatomen besitzen.

Eine naive Implementation dieser Suche würde über alle Fragmente und die dazugehörigen Linkatome gehen und sie mit der Anfrage vergleichen. Damit liegt sie in $O(N \cdot M)$ wobei N die Anzahl an Fragmenten und M die maximale Anzahl an Linkatomen eines Fragments ist. In den verwendeten BRICS Fragmenträumen ist $M \leq 5$ und kann im Verhältnis zur Anzahl der Fragmente vernachlässigt werden, wodurch sich eine lineare Komplexität ergibt. Trotz dieser moderaten Komplexitätsklasse stellte diese Berechnung eine nicht zu vernachlässigenden Zeitaufwand in Qsearch dar. Um die Suche zu beschleunigen, wurde ein Bitmapindex[93] eingesetzt, der die zu indizierenden Daten auf Bitmasken abbildet.

Um die Daten als Bitmaps zu kodieren, müssen sie in einer geeigneten Weise als Tabelle organisiert werden. Für dieses Problem wurde ein Tabellenschema entworfen, in dem jede Zeile ein Fragment darstellt und die Anzahl seiner jeweiligen Linkatomen in den Spalten beschrieben ist. (vgl. Tabelle 5.1) Dabei kodieren die einzelnen Spalten kumulativ die Anzahl der Linkatome eines bestimmten Typs. Das Tabellenschema erlaubt es, nach Fragmenten zu suchen, die mindestens eine bestimmte Kombination an Linkatomen besitzen. Bei der Suche wird die Anfrage mit dem gleichen Schema in eine Bitmaske

Algorithmus 5.1 Generieren der Menge aller Kombinationen von Linktypen N, die es einem Fragment g erlauben, ein Fragment f zu ersetzen. Die Eingabe m ist ein Vektor, der die Linktypen der Linkatome von f enthält. Die Instantiierung eines neuen Vektors wird durch eckige Klammern [] dargestellt. Die Funktion KOMP(l) liefert die zu l kompatiblen Linktypen als Vektor.

```
function GeneriereKompatibleLinkkombinationen(LinktypVektor m)
    k ← [ ] , a ← [ ] , r ← [ ]          ▷ Kompatibilitäten, Arbeitsvektor, Rückgabewert
    for l in m do
        k.ErweitereUm(komp(l))          ▷ Sammeln der kompatiblen Linktypen
        a.ErweitereUm(0)                ▷ Arbeitsvektor mit |m| mal 0 initialisieren
    end for
    while true do
        c ← [ ]        ▷ Zusammenstellen der aktuellen Kombination an Linkatomen
        for i ∈ (1, . . . , |m|) do
            c.ErweitereUm(k[i][a_i + 1])
        end for
        r.ErweitereUm(c)
        j ← |m|
        while a[j] = |k[j]| do                       ▷ Übertrag, wenn nötig
            a[j] ← 0
            j ← j − 1
            if j = 0 then        ▷ Abbruch, wenn alle Kombinationen erzeugt wurden
                return r
            end if
        end while
        a[j] ← a[j] + 1                    ▷ Erzeugen der nächsten Kombination
    end while
end function
```

	#Linkatome Typ l_1				#Linkatome Typ l_2				...	#Linkatome Typ l_k			
	≥ 1	≥ 2	...	$= m$	≥ 1	≥ 2	...	$= m$		≥ 1	≥ 2	...	$= m$
f_1	1	1			1					1			
f_2					1	1	1	1		1	1		
,,,													
f_n	1				1	1				1	1		

Tabelle 5.1.: Aufbau des Bitmapindexes. In den Spalten ist die Anzahl der Linktypen durch ein Bit repräsentiert. Voraussetzung ist die Nummerierung der Linktypen im Fragmentraums von l_1 bis l_k und die Nummerierung der Fragmente von f_1 bis f_n. Für ein Fragment f_i ist in der Zeile i das Bit in der Spalte $(t-1) \cdot m + d, d > 0$ gesetzt, wenn es mindestens d Links vom Typ l_t besitzt. Alle Fragmente besitzen mindestens 0 Linkatome aller Linktypen, daher sind diese Spalten ausgelassen. m ist die maximale Anzahl von Links des gleichen Typs, die ein Fragment im Fragmentraum hat.

umgewandelt, die für die Beschreibung der Fragmente verwendet wird und daraufhin zeilenweise mit dem Index verglichen. Wird eine Zeile gefunden, die mindestens die Bits der Anfrage gesetzt hat so wird das korrespondierende Fragment der Lösung hinzugefügt.

Algorithmus 5.2 Suche nach Fragmenten mit einer bestimmten Kombination von Linkatomen. Die Operation & ist die bitweise Verundung von zwei Bitmaps. Die Eingabe ist einerseits der generierte Index x ein Vektor mit Linktypen l als Anfrage.

```
function IndexSuche(Index x, LinktypVektor l)
    a ← [ ] , r ← [ ]                                      ▷ Anfrage, Rückgabewert
    for i in (1,. . . , x.AnzahlSpalten()) do
        a.ErweitereUm(0)                                   ▷ Initialisierung der Anfrage
    end for
    for i in (1, . . ., x.AnzahlLinktypen()) do            ▷ Konvertierung von l in Bitmaske
        for j in (1, . . ., x.MaxAnzahlLinks()) do
            if l.ZähleVorkommenVon(i)= j then
                a[(i − 1) · x.AnzahlLinktypen()+j] ← 1
            end if
        end for
    end for
    for i in (1, . . ., x.AnzahlZeilen()) do               ▷ Durchsuchen aller Zeilen
        if x.Zeile(i)&a = a then
            r.ErweitereUm(i)
        end if
    end for
    return r
end function
```

Um die Leistung des Indexes zu beurteilen wurde er in einem Experiment mit dem naiven Algorithmus verglichen. Dafür wurden Indizes für verschiedene Fragmenträume aufgebaut und jeweils tausend Anfragen durchgeführt, wobei zehn Anfragen zu einem Messpunkt zusammengefasst wurden, um die Zeitmessung genauer zu gestalten. Die Zeiten für die Ergebnisse sind in Abbildung 5.6.1 aufgeführt. Es lässt sich sehen, dass der Index die Rechenzeit um etwa vier Größenordnungen verringert. Darüber hinaus ist kann man erkennen, dass beide Methoden – sofern die wenigen Datenpunkte eine Interpolation erlauben – linear mit der Anzahl der Fragmente skalieren. Die Geschwindigkeitsvorteile durch den Index sind hauptsächlich technischer Natur und ergeben sich aus folgenden Punkten:

Cachelokalität Der Index speichert die Daten in einem zusammenhängenden Datenbereich. Dies erleichtert es dem Prozessor, die Daten im Rahmen des *Prefetching* im Voraus zu laden. Dadurch lassen sich Geschwindigkeitsvorteile von ungefähr zwei

Größenordnungen erreichen.[94]

Bitoperationen Die verwendeten Operationen basieren auf der Verknüpfung von Bitmasken. Diese Operationen sind für Prozessoren sehr schnell durchzuführen.

Komprimierung Durch die bit-basierte Kodierung der Daten wird die relevante Information auf weniger Speicherplatz abgebildet. Dadurch müssen insgesamt weniger Daten zwischen Hauptspeicher und Prozessor transportiert werden. Dies schlägt sich ebenfalls in den Ausführungszeiten nieder.

5.6.3. Zuweisung von Linkatomen eines Fragments zu den offenen Linkatomen des Restmoleküls

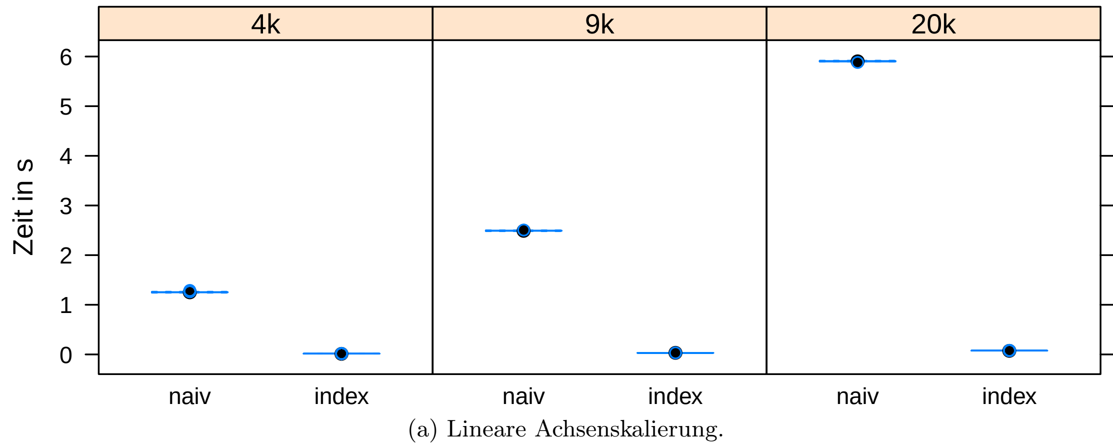
Nachdem in der Fragmentersetzung ein kompatibles Fragment g ausgewählt wurde, um f zu ersetzen, muss noch bestimmt werden, wie die Linkatome von g mit den Linkatomen des Restmoleküls L zu verknüpfen sind. Dafür werden zuerst alle möglichen Verbindungsmöglichkeiten generiert und dann eine davon zufällig ausgewählt. Die Generierung aller möglichen Verbindungsmöglichkeiten wird mithilfe des Algorithmus G („general permutation generator") von Knuth[95] durchgeführt. Der Algorithmus generiert für ein Tupel alle Permutationen der einzelnen Elemente, wobei sie in lexikographisch sortierter Reihenfolge aufgezählt werden. Dadurch können einfach Branch-and-Bound Beschränkung im Algorithmus implementiert werden (vgl. ebd.).

Die konkrete Implementierung ist in Algorithmus 5.3 beschrieben. Sie ordnet den offenen Linkatomen des Restmoleküls $l = (l_1, l_2, \ldots, l_n)$ die Linkatome $k = (k_1, k_2, \ldots, k_m)$ des Fragments g zu. Dabei wurde ein Branch-and-Bound Ansatz verwendet, um die Laufzeit zu begrenzen.[11]

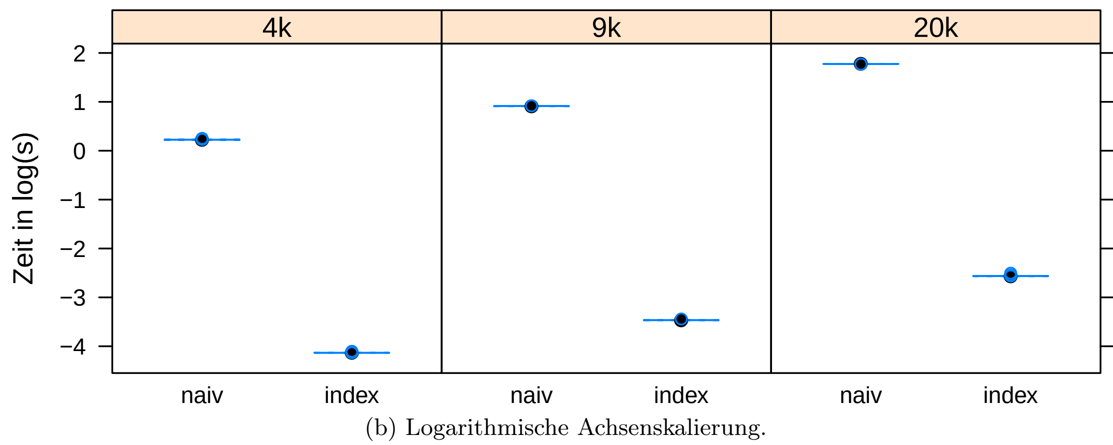
5.7. Fragmenterweiterung

Eine weiteres verbreitetes Anwendungsfeld für *de novo* Designmethoden ist die teilweise Modifikation von bestehenden Molekülen. Dies kann nötig werden, um die physikochemischen Eigenschaften eines Wirkstoffs zu verbessern, oder um geistiges Eigentum zu umgehen.[19] Mit einigen Modifikationen im Arbeitsablauf kann Qsearch auch für diese

[11] Während der Testläufe stellte sich heraus, dass die typische Problemgröße vernachlässigbar ist und die Verwendung des Branch-and-Bound Ansatzes keine messbaren Geschwindigkeitssteigerungen bringt. Praktisch alle Fragmente in den verwendeten Fragmenträumen haben zwei oder weniger Linkatome. Daher wurde dieser Ansatz trotz seiner Validität für einen besser wartbaren Programmcode verworfen, und die C++ Standardfunktion next_permutation verwendet.



(a) Lineare Achsenskalierung.



(b) Logarithmische Achsenskalierung.

Abbildung 5.6.1.: Box-Whisker Plots für den Vergleich des naiven Algorithmus gegen die Verwendung eines Indexes in zwei verschiedenen Skalierungen. Insgesamt wurden jeweils 1000 Anfragen durchgeführt. Ein Messpunkt beschreibt die Dauer von zehn Anfragen. Beim Vergleich ist zu beachten, dass sich die Größe der Datensätze von links nach rechts jeweils verdoppelt. Beide Algorithmen scheinen linear zu skalieren.

Algorithmus 5.3 Bestimmung aller Möglichkeiten zur Verbindung eines Fragments mit den Linkatomen k und dem Restmolekül mit den offenen Linkatomen l. Die Eingabe besteht aus zwei Vektoren von Linkatomen, die k beziehungsweise l repräsentieren, wobei $|k| \geq |l|$ ist. Der Algorithmus erzeugt Vektoren der Länge $|k|$ die jeweils eine Verbindungsmöglichkeit zwischen k und l beschreiben. Jeder Vektor a dieser Ausgabe besteht aus Elementen von l sowie und einer Anzahl von ε. Eine Stelle a_i des Vektors zeigt an, wie das Linkatom k_i zu verknüpfen ist. Steht an der Stelle ein Element von l, so repräsentiert dies eine Verknüpfung mit diesem Linkatom. Steht dort ein ε, so bleibt es unverknüpft. Bei der Sortierung ist ein ε kleiner als alle Werte, die l annehmen kann. Die Grundlage dieses Algorithmus, insbesondere das Erzeugen von Permutationen in lexikographisch aufsteigender Reihenfolge stammt von Knuth[95].

```
function ErzeugeKompatibleVerbindungen(LinktypVektor l, LinktypVektor k)
    r ← [ ] , a ← [ ]                                   ▷ Rückgabewert, Arbeitsvektor
    for i ∈ (1, . . . |k|) do              ▷ Generieren des Arbeitsvektors a der Länge |k|
        if i <= |l| then
            i.ErweitereUm(l[i])                         ▷ Setzen der |l| ersten Positionen
        else
            i.ErweitereUm(ε)                                          ▷ Auffüllen mit ε
        end if
    end for
    Sortiere(a)                        ▷ a muss die lexikographisch erste Permutation sein
    while true do                                                      ▷ Hauptschleife
        k ← |a| + 1
        for a_i in (a_n, . . . , a_1) do
            k ← k − 1      ▷ k + 1 beschreibt den Beginn des längsten kompatiblen Suffix
            if not kompatibel(a_i, l_i) then
                break
            end if
        end for
        if k = 1 then                                      ▷ Alle Stellen sind kompatibel
            r.ErweitereUm(a)
        end if
        if es gibt eine lexikogr. größere Permutation, die a_k verändert then
            a ← nächste lexikogr. größere Permutation, die a_k verändert
        else
            return r
        end if
    end while
end function
```

Problemstellung verwendet werden. Dabei wird ein vom Benutzer vorgegebenes Startfragment so mit den Fragmenten eines Fragmentraums erweitert, dass das vollständige Molekül einen Pharmakophor erfüllt. Diese abgewandelte Vorgehensweise wird „Fragmenterweiterung“ genannt und unterscheidet sich in folgenden Punkten:

- Das Startfragment darf nicht verändert werden. Das gilt sowohl für die fragmentbasierte Modifikation von Molekülen als auch für die Entdekorierungsschritte.
- In der Ausrichtungsphase darf sich das Startfragment nur wenig bewegen. Dies wird durch einen zusätzlichen Term in der Straffunktion erreicht, der für alle Schweratome des Startfragments die Distanz zu ihrem Ausgangspunkt bestraft. Die Strafe errechnet sich aus der Summe der quadrierten Entfernungen aller Startatome zu ihren jeweiligen Ursprungspositionen.

6. Ergebnisse und Diskussion

In diesem Kapitel wird eine Auswahl von Ergebnissen, die mit Qsearch in vier Anwendungsszenarien gewonnen wurden, vorgestellt und diskutiert. Um ein möglichst breites Bild über die Möglichkeiten und Einschränkungen der Methode zu gewinnen, wurden in den Szenarien der Pharmakophor, die Fragmenträume und der Modus (Fragmenterweiterung beziehungsweise freie Suche) variiert.

Das Kapitel beginnt mit einer Analyse der Fragmenträume, die als Eingabe verwendet wurden. In den darauf folgenden Anwendungsbeispielen wird mit COX-2 als Beispiel die Fähigkeit von Qsearch beleuchtet, auf Basis eines unspezifischen Pharmakophors Grundgerüste von bekannten aktiven Molekülen zu erzeugen. Dafür wurde zunächst ein targetspezifischer, und danach ein generischer Fragmentraum als Eingabe verwendet. In den weiteren Beispielen anhand eines spezifischen BCR-ABL Pharmakophors untersucht, wie sich die Grundgerüständerungen, die Qsearch vorschlägt, zu bekannten Änderungen aus der Literatur verhalten. Hierfür wurde ebenfalls mit einem allgemeinen Fragmentraum und einem targetspezifischen Fragmentraum gearbeitet, wobei letzterer im Rahmen eines Fragmenterweiterungsexperiments eingesetzt wurde.

Das Kapitel schließt mit einer Diskussion spezifischer Aspekte der Methode ab, wozu auch insbesondere ihre Einschränkungen gehören.

6.1. Analyse der verwendeten Fragmenträume

Vor der Präsentation der durchgeführten Experimente werden die verwendeten Fragmenträume analysiert, da sie einen maßgeblichen Einfluss auf die Qualität der Ergebnisse haben. Dafür werden sie in Bezug auf verschiedene Eigenschaften untersucht und ihre Eignung für die Anwendung in Qsearch beurteilt. Ein besonderes Augenmerk gilt dabei der Diversität der Fragmente in Bezug auf ihre Geometrie, Topologie und Chemie.

Die gewonnenen Erkenntnisse wurden in die Erstellung eines neuen Fragmentraums umgesetzt, mit dem angestrebt wurde, die Diversität der Fragmente auf den eben genannten Felder zu erhöhen. Dazu wurde der BRICS_4k Raum mit Fragmenten des BRICS_20k Raums angereichert, die bestimmten Kriterien entsprechen: Die geometri-

sche Diversität wurde dadurch erhöht, dass Fragmente, die ein nicht-planares Ringsystem besitzen, hinzugefügt wurden, und die topologische Diversität dadurch, dass Fragmente hinzugefügt wurden, die mehr als ein Linkatom aufweisen. Der daraus entstandene Fragmentraum wird im folgenden als „BRICS_4k+“ Fragmentraum bezeichnet.

6.1.1. Geometrische Diversität

Bedeutung Da die *molekulare Erkennung* zu einem großen Teil über die geometrischen Eigenschaften von Molekülen bestimmt wird, ist im *de novo* Design eine hohe geometrische Diversität bei den verwendeten Fragmente wünschenswert um ein breites Spektrum an molekularen Formen abzudecken.[96]

Die geometrische Diversität von Fragmenten wird über zwei Parameter beeinflusst. Der erste ist die Flexibilität innerhalb der Fragmente, die durch das Gegeneinanderverdrehen von Molekülteilen an rotierbaren Bindungen entsteht. Je mehr aufeinander folgende rotierbare Bindungen in einem Fragment existieren, desto größer ist die Spannbreite an Formen, die von ihm angenommen werden können. Zu diesem Parameter zählt auch das Wechseln von Ringkonformationen an flexiblen Ringen.[1]

Der zweite Parameter ist die geometrische Lage der Linkatome zueinander. Sie ist unter dem Begriff *Exitvektor*[40] bekannt, und bestimmt die Formen, die durch das reine Verknüpfen von Fragmenten erhalten werden kann. Befinden sich die Linkatome an starren molekularen Strukturen, so kann durch die Verknüpfung mehrerer solcher Fragmente eine große Vielfalt an starren molekularen Formen gewonnen werden.

Von den beiden Möglichkeiten, die geometrische Variabilität in Fragmenträumen zu beeinflussen, ist im Kontext der Wirkstoffforschung die Verknüpfung von starren Fragmenten zu bevorzugen, da eine erhöhte Flexibilität über den Entropieterm das chemische Gleichgewicht von Protein-Ligand Komplexen auf die Seite der ungebundenen Edukte verschiebt.[91] Die Bindungsaffinität von Wirkstoffen wird also durch eine höhere Flexibilität innerhalb von Liganden erniedrigt.

Statistische Analyse Im Falle der BRICS Räume lässt sich feststellen, dass die geometrische Diversität der Fragmente in Bezug auf die Anordnung der Exitvektoren nicht stark ausgeprägt ist. In Abbildung 6.1.1 ist erkenntlich, dass ein Großteil der in ihnen enthaltenen Fragmente planar oder nahezu planar ist. Dies gilt für circa 70% der Moleküle in den Standard-BRICS-Räumen und circa 55% der Moleküle in dem angereicherten BRICS-Raum, und ist eine direkte Konsequenz der Methode, mit der die Fragmenträume

[1] Qsearch berücksichtigt das Wechseln von Ringkonformationen nicht explizit. Stattdessen muss für jede Ringkonformation ein Fragment in der Eingabe vorhanden sein.

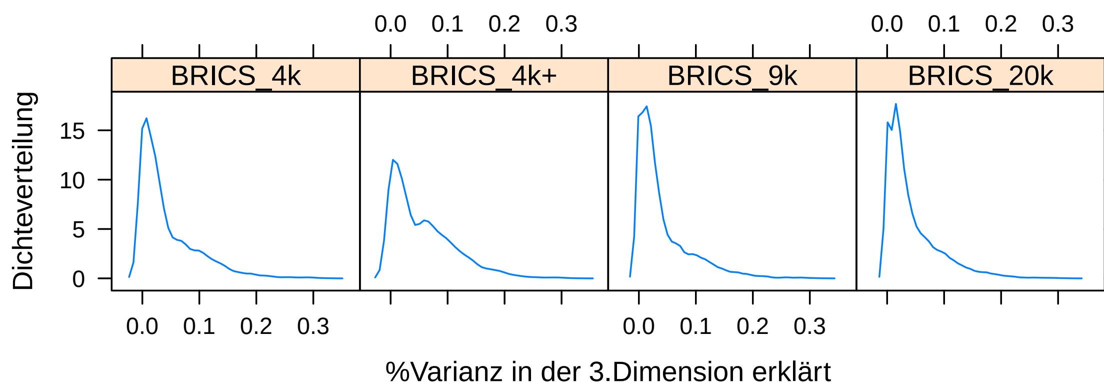


Abbildung 6.1.1.: Kerndichteschätzung für die Planarität der verwendeten Fragmente. Aufgetragen ist der Anteil der Varianz der Koordinaten, die nach einer Hauptkomponentenanalyse durch die dritte Komponente erklärt wird. Der größte Anteil der Fragmente besitzt planare Ringsysteme, was in den Maxima der Standard BRICS-Räume erkennbar ist. Gut zu sehen ist ebenfalls, dass im angereicherten BRICS_4k+ Raum der Anteil der Fragmente mit planaren Ringsystemen niedriger ist, und dafür ein zweites Maximum – vermutlich durch nicht-planare Ringsysteme – hinzugekommen ist.

erstellt worden sind. Die BRICS Räume sind auf Ähnlichkeit zu Wirkstoffbibliotheken optimiert, wozu auch eine Präferenz von aromatischen Ringsystemen gehört, die wiederum ein weit verbreitetes Strukturmotiv in Wirkstoffen sind. Folglich sind in den Räumen aromatische Ringsysteme ebenfalls stark vertreten.

Konsequenzen In den von Qsearch erzeugten Ergebnissen lässt sich sehen, dass viele Moleküle hauptsächlich aus planaren Untereinheiten bestehen, und die geometrische Vielfalt größtenteils durch rotierbare Bindungen zwischen ihnen zustande kommt. Diese Eigenschaften sind insbesondere bei den Ergebnissen für den BCR-ABL Pharmakophor ausgeprägt.

Diskussion Das Entwerfen von Fragment- und Molekülbibliotheken, die ein breites Spektrum an Geometrien abdecken, ist sowohl für virtuelle Fragmenträume als auch für experimentelle Fragmentbibliotheken von großer Bedeutung.[96] Bisher fokussieren sich die aktuellen Methoden für das Bibliotheksdesign allerdings auf die chemische und nicht auf die geometrische Diversität. (vgl. ebd.) Von einer Erhöhung der geometrischen Diversität in der Eingabe könnten die Ergebnisse von Qsearch substanziell profitieren, da sie ein breiteres Spektrum an molekularen Formen abdecken würden. Allerdings würde

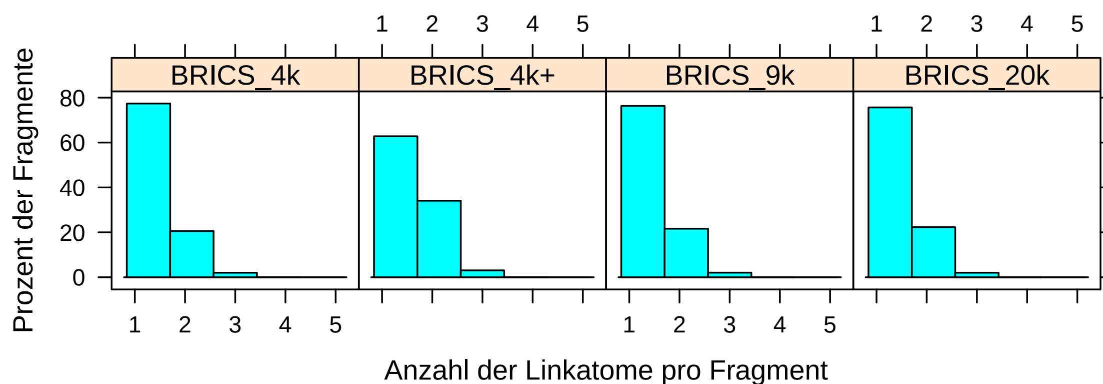


Abbildung 6.1.2.: Analyse der topologischen Diversität der BRICS Räume. Aufgetragen sind Histogramme über die Anzahl an Linkatomen pro Fragment in den vier Beispielfragmenträumen. Es ist ersichtlich, dass die klassischen BRICS Räume hauptsächlich aus terminalen Fragmenten bestehen. Der angereicherte BRICS_4k+ Raum enthält mehr Fragmente mit zwei Linkatomen, aber trotzdem nur minimal mehr Fragmente mit drei Linkatomen.

durch diesen Schritt auch die Wirkstoffähnlichkeit der Fragmenträume verringert werden, da sie weniger aromatische Systeme enthalten würden, die bei Wirkstoffen häufig anzutreffen sind.

6.1.2. Topologische Diversität

Bedeutung Eines der Hauptanwendungsfelder für das *de novo* Design ist die Suche nach neuartigen molekularen Grundgerüsten. Trotz des Fehlens einer allgemeingültigen Definition des Begriffs „Grundgerüst“, kann die Topologie der molekularen Strukturformeln als ein bestimmendes Element im Grundgerüstkonzept angesehen werden. (vgl. Abschnitt 2.4) Folglich ist eine große Diversität der Fragmente in Bezug auf die Topologie eine der Grundlagen für den Zugang zu möglichst diversen Leitstrukturen.

Statistische Analyse In Abbildung 6.1.2 ist die Verteilung der Fragmente in Hinblick auf die Anzahl der Linkatome in den verwendeten Fragmenträumen aufgetragen. Dabei ist klar ersichtlich, dass die Verteilung in sämtlichen Fragmenträumen stark auf terminale Fragmente mit nur einem Linkatom konzentriert ist. Nur ungefähr 20% der Fragmente weisen zwei oder mehr Linkatome auf. In dem angereicherten BRICS_4k+ Raum finden sich mehr nicht-terminale Fragmente, allerdings besitzen auch hier ca. 60% der Fragmente nur ein Linkatom.

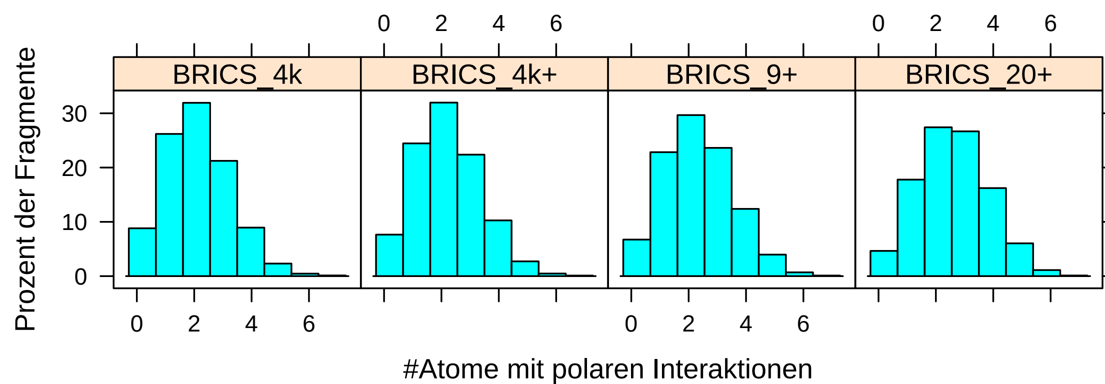
Konsequenzen Die Konsequenzen der präsentierten Topologieverteilung in den BRICS-Räumen spiegeln sich in den präsentierten Ergebnissen wider. So hat die hohe Anzahl der Fragmente mit einem Linkatom unter anderem Einfluss auf die Ergebnisse für den COX-2 Pharmakophor, wo fast alle Lösungen aus zwei terminalen Fragmenten bestehen. Für den BCR-ABL Pharmakophor konnten nur wenige Moleküle gefunden, die eine nicht-lineare Topologie aufweisen, in denen also Fragmente mit mehr als zwei Linkatomen verbaut wurden.

Diskussion Die Anzahl der Linkatome ist kein allumfassendes Maß zur Beschreibung der topologischen Diversität in Fragmenträumen. Es ist denkbar, dass Fragmente nur ein oder zwei Linkatome besitzen, aber durch größere interne Verzweigungen komplexe Topologien aufweisen. Um die topologischen Diversität in Fragmenträumen beurteilen zu können, müsste ein Maß sowohl für die interne Topologie von Fragmenten als auch für die Topologien, die durch das Verknüpfen von Fragmenten entstehen können, erarbeitet werden. Allerdings ist auch ohne dieses Maß zu vermuten, dass in den BRICS-Räumen die Anzahl an nicht-terminalen Fragmenten für eine breite Abdeckung an diversen Grundgerüsten zu niedrig ist. Wie jedoch das optimale Verhältnis der Anzahl an Linkatomen in den Fragmenten aussieht, muss Teil weiterer Forschungsarbeiten sein.

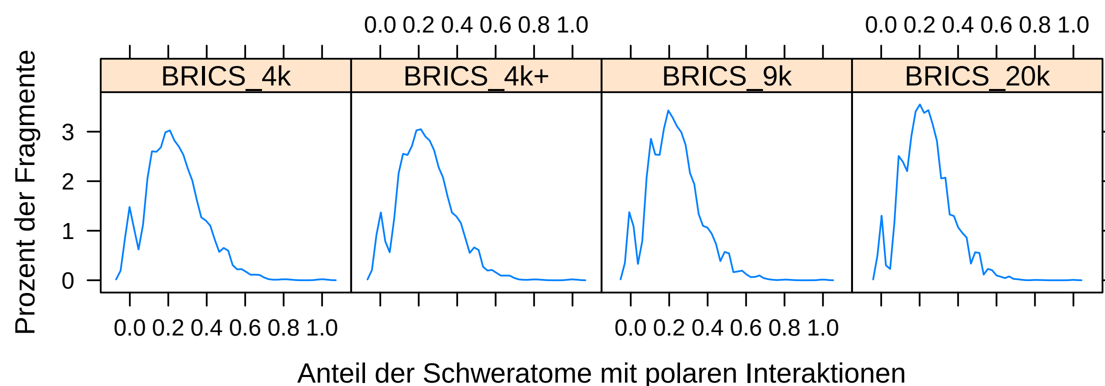
6.1.3. Chemische Diversität I: Hydrophobizität

Bedeutung Wie in Abschnitt 6.5.1 beschrieben, ist die Hydrophobizität einer der wichtigsten Faktoren um die Eignung einer Substanz als Leitstruktur zu beurteilen. Dabei gibt es mehrere Möglichkeiten um die Hydrophobizität von Molekülen zu beschreiben. In den „Lipinski rule of 5“ Regeln[97], die als Faustregel für die orale Bioverfügbarkeit von Wirkstoffen verwendet werden, wird sowohl der Oktanol-Wasser Partitionskoeffizient als auch die Anzahl an Donor- und Akzeptoratomen als Maß verwendet. Das Hydrophobizitätskonzept von Qsearch basiert auf letzterem, weswegen die Hydrophobizität für die verwendeten Fragmente im folgenden ebenfalls durch die Anzahl der Atome, die als Donor- oder Akzeptor klassifiziert werden, beschrieben wird.

Statistische Analyse In Abbildung 6.1.3 ist die Anzahl der polaren Atome pro Fragment und das Verhältnis von polaren zu nicht polaren Schweratomen in den verwendeten Fragmenträumen graphisch aufgetragen. Es fällt auf, dass ein Großteil der Fragmente zwei oder mehr polare Atome beinhaltet. Darüber hinaus weisen die meisten Fragmente ein Verhältnis von polaren zu unpolaren Schweratomen von 1:4 oder mehr auf. Der Median für den Anteil an Heteroatomen in allen Datensätze in diesem Fall jeweils 0,22. Daraus



(a) Anzahl der polaren Schweratome in den Fragmenten.



(b) Anteil der polaren Atome an den gesamten Schweratomen in den Fragmenten.

Abbildung 6.1.3.: Polarität der Fragmente in den BRICS Fragmenträumen. Atome werden als polar angesehen, wenn sie im Interaktionsmodell entweder als Akzeptor oder als Donor behandelt werden.

erschließt sich, dass mehr als jedes zweite Fragment zu 20% aus polaren Schweratomen besteht.

Konsequenzen Eine der auffälligsten Konsequenzen für Qsearch ist die hohe Polarität der Moleküle, die als Ergebnis geliefert werden. Da in den Fragmenten eine hohe Anzahl an polaren Gruppen vorhanden ist, finden sich diese – sofern sie nicht durch den in Abschnitt 5.4.2 beschriebenen Entdekorierungsschritt entfernt werden – in den Ergebnismolekülen wieder. Darüber hinaus ist die Verwendung von hydrophoben Pharmakophormerkmalen mit den BRICS Fragmenträumen stark beeinträchtigt, da es kaum Fragmente ohne polare Atome gibt.

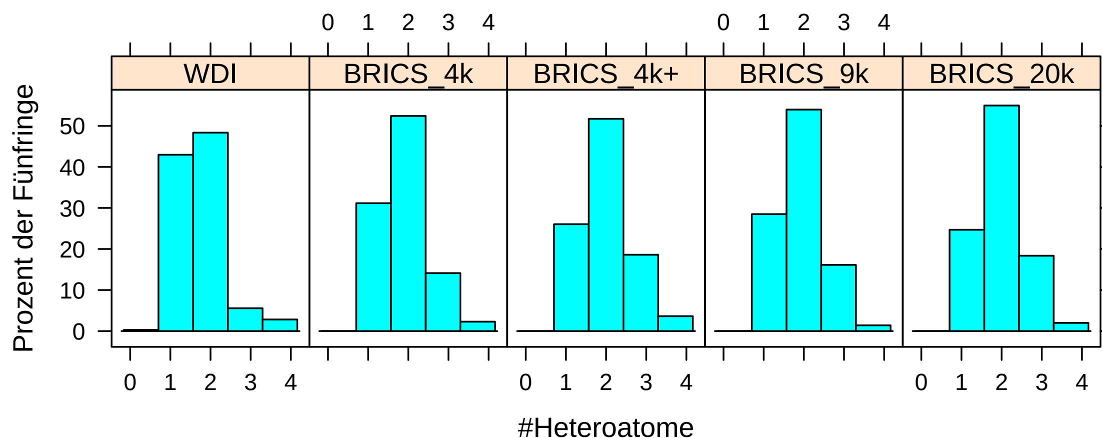
Diskussion Die hohe Zahl an polaren Fragmenten in den BRICS Räumen lässt sich einfach auf die Datensätze zurückführen, die benutzt wurden, um die Fragmente zu generieren. Zum einen sind die verwendeten Moleküle voll optimierte Wirkstoffe, so dass bereits polare Gruppen für die Erhöhung der Löslichkeit eingeführt wurden. Zum anderen wurden aromatische Heterozyklen im Fragmentauswahlprozess überrepräsentiert, was im nächsten Abschnitt gezeigt wird.

Die direkten Konsequenzen durch polare Fragmente für Qsearch halten sich allerdings in Grenzen. Sie stellen vor allen Dingen ein großes Hindernis bei der Akzeptanz der Lösungen bei den Endanwendern dar, die polare Gruppen nur bei entsprechenden Pharmakophormerkmalen erwarten. Vielmehr ist es für den Algorithmus eher förderlich, Fragmenten mit möglichst vielen funktionellen Gruppen zu haben, von denen die nicht benötigten im Entdekorierungsschritt entfernt werden. So kann Qsearch auf einen größeren Raum an Arrangements von funktionellen Gruppen zurückgreifen.

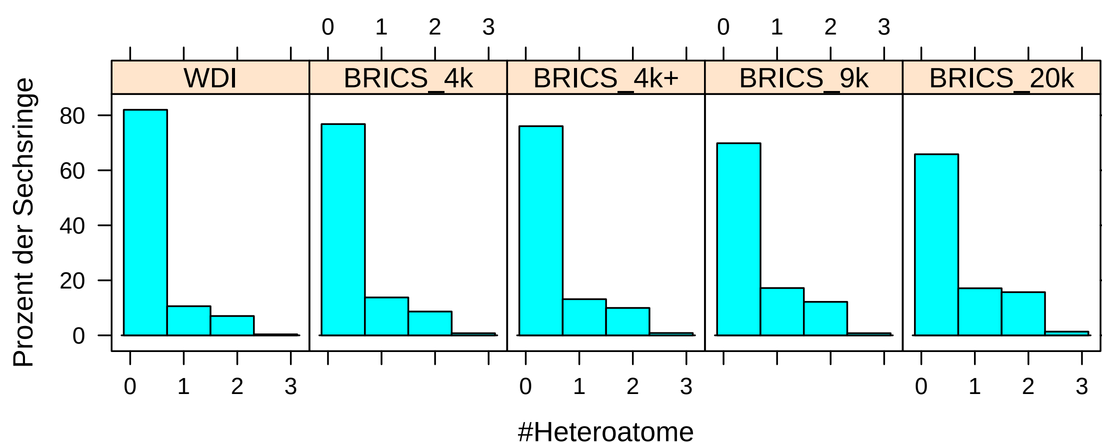
6.1.4. Chemische Diversität II: aromatische Heterozyklen

Bedeutung Ringsysteme finden in Wirkstoffen sowohl als Strukturmotiv als auch als funktionelle Gruppe breite Anwendung, wobei sie entweder dazu dienen können, interagierende Gruppen in einem bestimmten Abstand zueinander zu fixieren, oder selbst Interaktionen auszubilden. Dabei haben aromatische Heterozyklen regulären Ringen gegenüber den Vorteil, dass sie in der Regel keine Stereozentren tragen. Aromatische Heterozyklen liegen klassischerweise als Ringe mit fünf oder sechs Atomen vor, wobei die Sechsringe im WDI relativ häufiger als Fünfringe vorkommen.

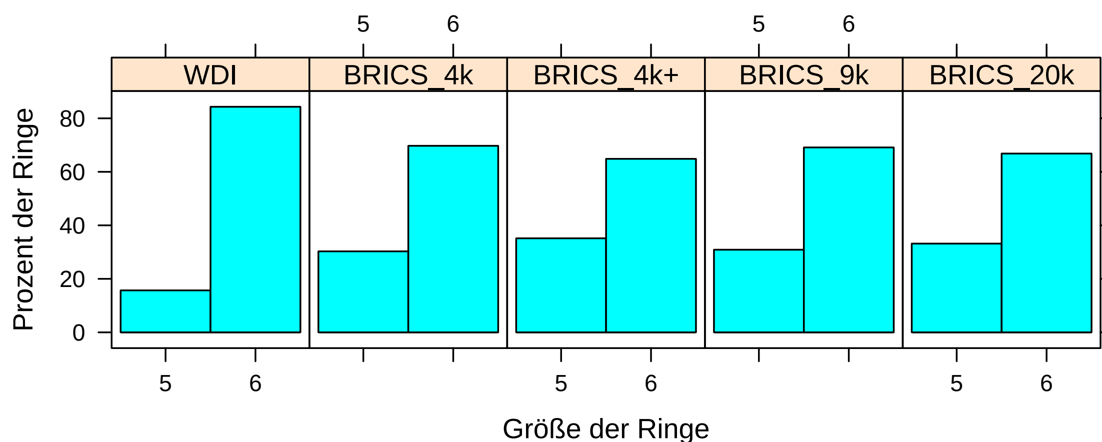
Statistische Analyse Die Ergebnisse der Analyse sind in Abbildung 6.1.4 zusammengefasst. Es ist ersichtlich, dass bei den Fünfringen der verwendeten Fragmenträume eine



(a) Verteilung von Heteroatomen in aromatischen Fünfringen. In den BRICS-Räumen gibt es bedeutend mehr Ringe mit zwei Heteroatomen als im WDI.



(b) Verteilung von Heteroatomen in aromatischen Sechsringen. Ringe mit mehr als drei Heteroatomen sind ausgelassen, da sie nur selten vorkommen. Die Verteilung ist annähernd identisch mit der im WDI.



(c) Verteilung der Ringe in dem WDI und den Fragmenträumen.

Abbildung 6.1.4.: Analyse der Eigenschaften von Heterozyklen in verschiedenen Fragmenträumen.

starke Überrepräsentation von Ringen mit zwei oder mehr Heteroatomen vorhanden ist, die sich so in den Ergebnissen wiederfindet. Im Vergleich dazu ist die Verteilung bei den Sechsringen annähernd gleich.

Im Zuge dieser Analyse wurde zusätzlich die Verteilung von Fünf- und Sechsringen in den Fragmenten erforscht, wobei sich herausstellte, dass innerhalb dieser Fragmente die Fünfringe viel stärker repräsentiert sind als im WDI. (Abbildung 6.1.4) Während dort 15,7% der Ringe fünf Atome haben, sind es in den BRICS-Räumen über 30%.

Konsequenzen Für Qsearch ergeben sich aus den gefundenen Verteilungen bei den Heterozyklen in den Fragmenträumen zwei Folgen. Zum einen entspricht das Verhältnis von Fünf- und Sechsringen in den Lösungen nicht dem, das man bei Wirkstoffen erwartet (vgl. WDI-Werte in Abbildung 6.1.4). Zum anderen ist die Anzahl an Heteroatomen in den Fünfringen erhöht.

Diskussion Anders als bei der Verringerung der Heteroatomanzahl in der Entdekorierungsphase ist eine Korrektur des Verhältnisses von Fünf- zu Sechsringen nicht ohne weiteres möglich. Die Ringe sind geometrisch nicht äquivalent und ein verschobenes Verhältnis in der Eingabe könnte von den Benutzern beabsichtigt sein, so dass eine Korrektur nicht durchgeführt werden sollte. Für einen Ausgleich der Verhältnisse sind weitere Arbeiten nötig, die die Gründe für ihre Verschiebung beleuchten.

Die Überrepräsentierung der Fünfringe resultiert in einem weiteren Problem. Während Sechsringe leicht durch Benzenringe ohne Heteroatome ersetzt werden können, so muss in aromatischen Fünfringen stets mindestens ein Heteroatom vorhanden sein, um die Aromatizität zu gewährleisten, was die Anzahl der Nichtkohlenstoffatome zu Lasten der Benutzerakzeptanz erhöht.

In die Erstellung des BRICS_4k+ Raumes floss bisher noch keine Bemühung, die erwähnten Probleme auszugleichen. Es ist jedoch davon auszugehen, dass die Qualität und vor allen Dingen die Nutzerakzeptanz der Ergebnisse über eine solchen Schritt zu steigern ist.

6.2. COX-2 Inhibitoren

Nichtsteroidale Antirheumatika und die Cyclooxygenase-2

Die Cyclooxygenase ist ein Enzym, das die Entstehung von Prostaglandinen aus Arachidonsäure katalysiert. Da Prostaglandine im Körper als Schmerzmediatoren wirken, hat

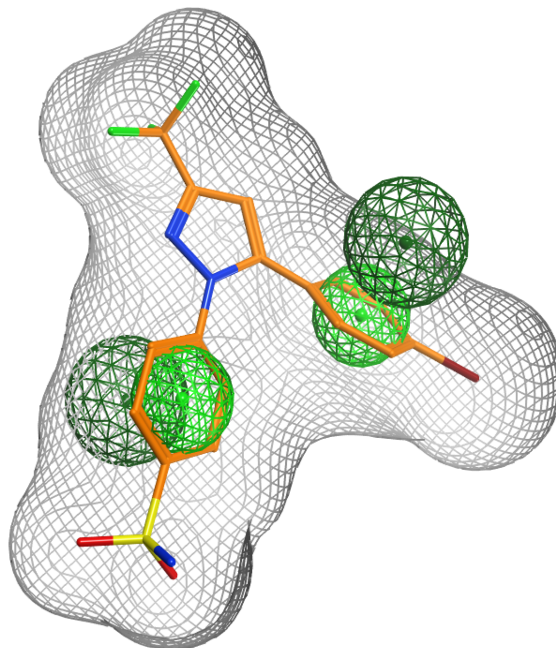


Abbildung 6.2.1.: Der verwendete COX-2 Pharmakophor. Hellgrün: aromatische Zentren, dunkelgrün: aromatische Richtungen. grau: erlaubte Form. Das dargestellte Molekül ist ein Derivat des COX-2 Inhibitors Celecoxib aus der PDB Struktur 6cox.

ihre Blockade eine Herabsetzung der Schmerzempfindlichkeit zur Folge.[98] Darüber hinaus fungieren Prostaglandine als Botenstoff, der Entzündungsreaktionen hervorruft und die Aggregation von Thrombozyten unterdrückt. Folglich werden Cyclooxygenasehemmer als Analgetika, Antiphlogistika und Thrombozytenaggregationshemmer eingesetzt. Entzündungshemmende Schmerzmittel, die wie die COX-Inhibitoren kein steroidales Grundgerüst besitzen, werden auch als nichtsteroidale Antirheumatika (NSAR) bezeichnet.[99]

Die Cyclooxygenase liegt in den zwei strukturellen Isoformen COX-1 und COX-2 vor. Die NSAR der ersten Generation hemmen nicht-selektiv beide Isoformen, aber durch die Aufklärung der strukturellen Unterschiede zwischen ihnen konnte sich die Forschung auf die Entwicklung von selektiven COX-2 Inhibitoren konzentrieren. Diese Isoform hat einen weitaus größeren Anteil an der Schmerzmediation, und zeigt bei einer Inhibition im Vergleich weniger Nebenwirkungen.[1]

COX-2 Pharmakophor

Die Bindetasche von COX-2 ist weitgehend hydrophob und die *molekulare Erkennung* von Liganden wird hauptsächlich durch ihre Form bestimmt.[54] Um zu untersuchen, ob

die in Qsearch implementierte Methodik Topologien bekannter Inhibitoren reproduzieren kann, wurde auf Basis dieser Tasche ein Pharmakophor erstellt, der bereits strukturelle Eigenschaften über aromatische Merkmale vorgibt. So konnten die Lösungen einfach mit den Grundgerüsten von bekannten COX-2 Inhibitoren verglichen werden. Der erstellte Pharmakophor ist in Abbildung 6.2.1 dargestellt. Er besteht aus zwei gerichteten aromatischen Pharmakophormerkmalen und einem Volumen, das auf Basis des Celecoxib-Analogs in der PDB Struktur 6cox[100] erstellt wurde. Es beschreibt die Tasche, in der die Inhibitoren binden, und wurde durch das Platzieren von Kugeln mit dem Radius 1,8Å auf allen Schweratomen des Liganden definiert. Ein vergleichbarer Ansatz wurde in der Validierung des Programms SkelGen[54] gewählt, wo die Pharmakophordefinition ähnlich unspezifisch war.

Generierung eines targetspezifischen COX-2 Fragmentraums

Für das erste Experiment mit dem COX-2 Pharmakophor wurde ein targetspezifischer Fragmentraum für COX-2 Inhibitoren generiert. Dafür wurde ein Datensatz mit 128 Molekülen, die eine Aktivität gegenüber COX-2 aufweisen, von Stahl und Rarey[101] übernommen und mithilfe der BRICS Schnittregeln zu insgesamt 414 Fragmenten zerlegt. In diesem speziellen Kontext bietet der Fragmentraum neben funktionellen Gruppen, die für COX-2 Inhibition typisch sind, den zusätzlichen Nutzen, dass er die Vergleichbarkeit der generierten Ergebnisse mit bekannten Inhibitoren vereinfacht, da der Suchraum kleiner ist, und die vorkommenden Strukturmotive einfach denen der aktiven Molekülen zuzuordnen sind.

Durchführung und Ergebnisse

Mit dem targetspezifischen Fragmentraum und dem oben beschriebenen Pharmakophor als Eingabe wurde das Programm 8 Minuten lang[2] ausgeführt. Als Besonderheit wurde in diesen Läufen der Entdekorierungsschritt ausgelassen, um die speziellen Eigenschaften des Fragmentraums zu bewahren. Dabei wurden genau 100 Moleküle, die dem Pharmakophor genügen, erzeugt. Die generierten Moleküle wurden manuell den verallgemeinerten Grundgerüsten aus der SkelGen Validierung von Stahl et al.[54] zugeteilt. Die Ergebnisse sind in Tabelle 6.1 aufgeführt, wo zu jedem Grundgerüst mindestens eine entsprechende Lösung abgebildet ist.

In einem zweiten Experiment diente statt des targetspezifischen Fragmentraums der BRICS_4k+ Fragmentraum als Eingabe. Das Programm wurde auf der gleichen Rech-

[2]Laptop mit Intel Core2 Duo mit 2GHz, 4GB Speicher, OpenSuse 11.3.

Grundgerüst	Qsearch	Bekannter Inhibitor	Grundgerüst	Qsearch	Bekannter Inhibitor

Tabelle 6.1.: Auswahl von Molekülen, die im COX-2 Beispiel aus dem targetspezifischen Fragmentraum generiert wurden. Erste Spalte: verallgemeinertes Grundgerüst, mittlere Spalte: generierte Lösung, rechte Spalte: bekanntes aktives Molekül. In blau (4. Zeile, rechte Seite) ist eine Lösung mit einem neuartigen Ringschluss hervorgehoben.

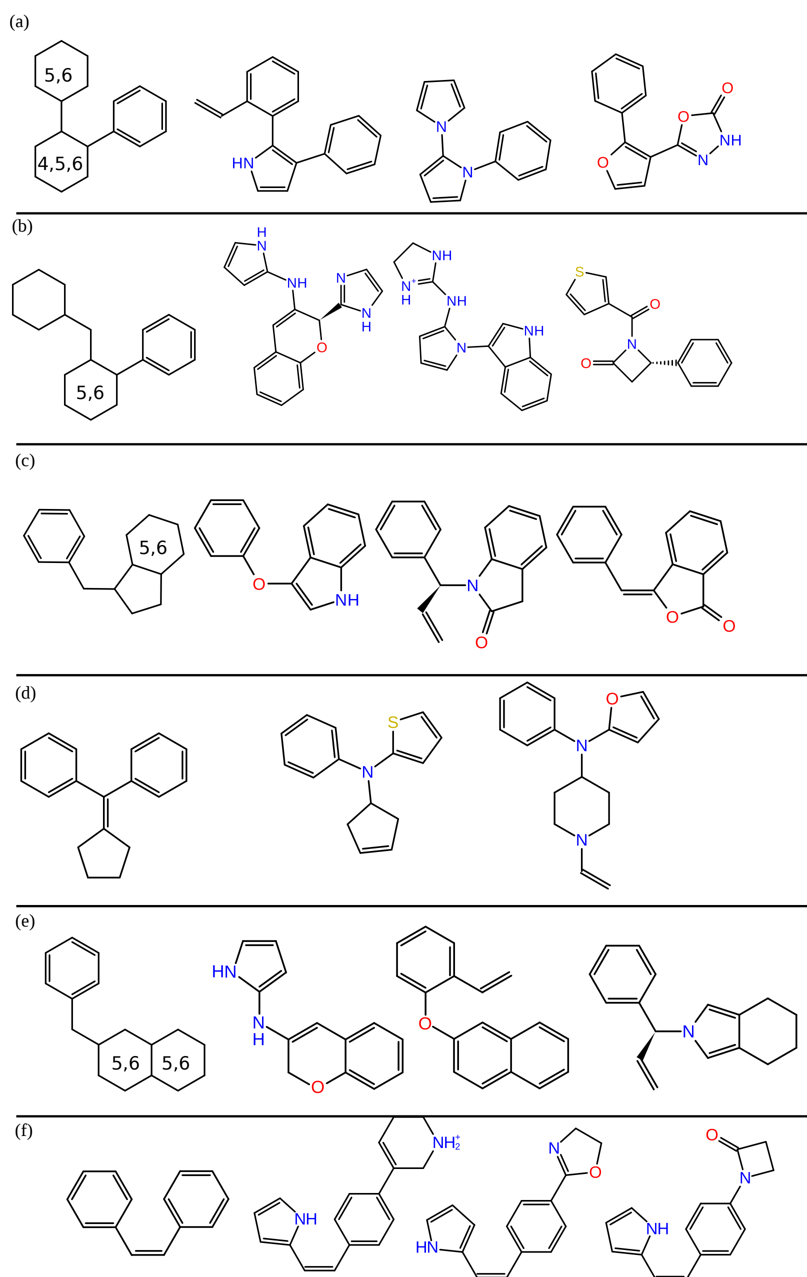


Tabelle 6.2.: Bekannte Grundgerüste von COX-2 Inhibitoren neben von Qsearch erstellten Lösungen. In der ersten Spalte sind generalisierte Darstellung der molekularen Grundgerüste abgebildet. Die Zahlen in den Ringen repräsentieren eine variable Anzahl an Atomen. Gezeichnete Einfachbindungen können sowohl Einfach- als auch Doppelbindungen repräsentieren. Daneben sind von Qsearch generierte Lösungen zu sehen.

nerumgebung ausgeführt, allerdings aufgrund des größeren Suchraums mit 16 Minuten doppelt so lange, wobei circa 500 Lösungen erhalten wurden. Die Lösungen wurden ebenfalls manuell den bekannten Grundgerüsten zugeordnet und in Tabelle 6.2 mit ihren korrespondierenden Grundgerüsten aufgetragen. Genau wie im ersten Beispiel konnten Repräsentanten der bekannten Molekülgerüste von COX-2 Inhibitoren unter den generierten Lösungen gefunden werden. Die einzige Ausnahme bildet die Klasse (d) in der o.g. Tabelle. Diese Klasse zeichnet sich durch ein zentrales planares sp^2-hybridisierte Kohlenstoffatom aus, an dem zwei aromatische Ringe und ein aliphatischer Ring gebunden sind. Das zentrale Kohlenstoffatom konnte in keinem Ergebnismolekül gefunden werden. Stattdessen wurden Lösungen erzeugt, in denen ein geometrisch äquivalentes Stickstoffatom diese Position ausfüllt.

Zusätzlich zu den Grundgerüsten aus der SkelGen Validierung, die aus dem targetspezifischen Fragmentraum reproduziert werden konnten, wurden Moleküle erzeugt, die ein cis-Stilben Motiv aufweisen. (Klasse (f) in der Tabelle 6.2) Dieses Motiv war in dem targetspezifischen Fragmentraum nicht enthalten, repräsentiert jedoch ebenfalls eine Klasse von aktiven Molekülen.[102] Im Vergleich zu den klassischen cis-Stilbenen besitzen die erzeugten Moleküle an ihren Phenylringen jeweils einen zusätzlichen Ring in para-Stellung, der von Qsearch angefügt wurde, um das geforderte Minimalvolumen zu erreichen.

Interpretation

In dem ersten Beispiel ließ sich zeigen, dass für jede der Grundgerüstklassen der COX-2 Inhibitoren aus der SkelGen Validierung Lösungen generiert werden konnten. Die in Tabelle 6.1 abgebildeten bekannten aktiven Moleküle konnten entweder exakt oder annähernd exakt reproduziert werden. Dazu gehören zum Beispiel die Lösungen mit dem Cyclopentanring oder dem Dihydrofuran-2-on Rest. (1. bzw. 2. Zeile rechts) Darüber hinaus ließ sich eine Lösung finden, die die terminale Doppelbindung des Grundgerüsts ohne den angeschlossenen fünfgliedrigen Ring aufweist (3. Zeile rechts). Sie ist vergleichbar zu dem daneben abgebildeten offenkettigen Vorläufer der Struktur mit dem Dihydrofuran-2-on Ring. Die in der Tabelle blau dargestellte Struktur mit dem neuen Ringschluss entspricht keinem der angegebenen Grundgerüste und lässt sich als neuartig bezeichnen. Auch die dreidimensionalen Überlagerungen mit dem Molekül, das dem Pharmakophor als Grundlage diente, zeigen eine gute Ähnlichkeit. (Abbildung 6.2.2)

In dem zweiten Experiment konnten in den Lösungen ebenfalls Repräsentanten aller Grundgerüstklassen gefunden werden. Im Vergleich zu den Ergebnissen, die mit dem targetspezifischen Fragmentraum generiert wurden, waren die Moleküle wie erwartet unähnlicher zu den bekannten COX-2 Inhibitoren. Dies ist hauptsächlich auf die Dekoration

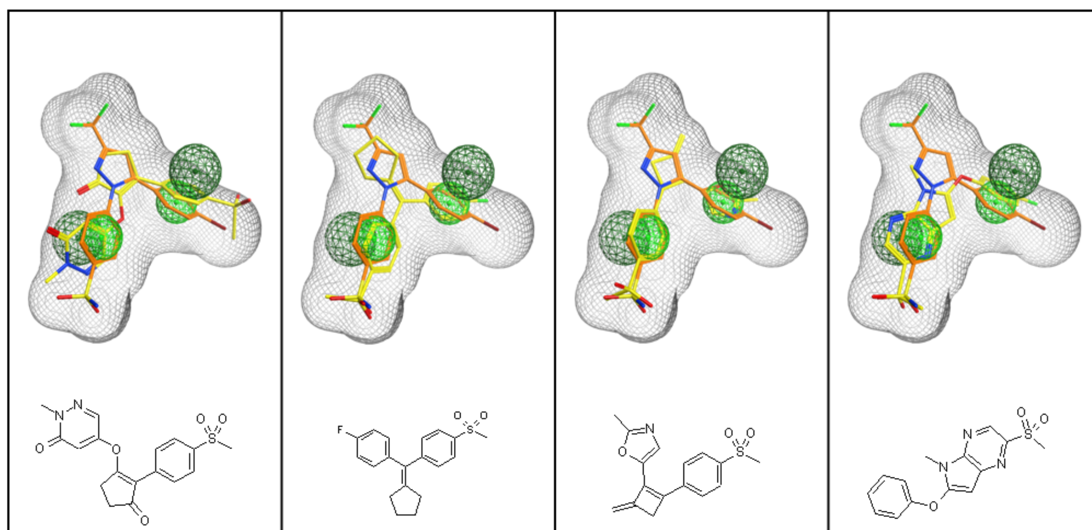


Abbildung 6.2.2.: Vier Qsearch Ergebnismoleküle (gelb) überlagert mit dem Pharmakophor und dem Molekül, mit dem die Anfrage generiert worden ist (orange). Unten ist die 2D-Strukturformel der jeweiligen Moleküle zu sehen.

der Fragmente und nicht auf die Grundgerüste zurückzuführen, da die verwendeten Fragmente nicht die für COX-2 typischen funktionellen Gruppen tragen. Auffällig ist ebenfalls, dass die Lösungen trotz des Entdekorierungsschritts teilweise sehr polar sind und viele Fünfringe enthalten. Mögliche Ursachen für diese Beobachtungen werden in Abschnitt 6.1.4 besprochen. Anhand dieses Beispiels lässt sich sehen, wie die Verwendung eines generischen Fragmentraums die Generierung von neuen Grundgerüsten (in diesem Fall das der cis-Stilbene) ermöglichen kann, die bei der Verwendung eines targetspezifischen Fragmentraums nicht gefunden werden konnten.

Für die Grundgerüstklasse, in der der zentrale sp^2-hybridisierte Kohlenstoff als zentrales Verzweigungsatom dient (Klasse (d) in Tabelle 6.2), wurden nur Moleküle mit planaren Stickstoffatomen generiert. Hier zeigt sich, wie die Ergebnisse von Qsearch als Grundgerüste gelesen werden sollten. Die relevante Information in diesen Molekülen ist, dass ein zentrales planares Verbindungsatom die aromatischen Ringe so positionieren kann, dass sie den Pharmakophor geometrisch erfüllen. Spezifischere Informationen, wie zum Beispiel das Element des Atoms sind in diesem Fall weniger interessant, vor allen Dingen, da es in diesem Fall nicht für die Erfüllung der elektronischen Pharmakophormerkmale benötigt wird.

Um im zweiten Versuch bei den Ergebnissen eine Diversität zu erreichen, die mit der

des ersten Versuchs vergleichbar ist, musste die Laufzeit von 8 auf 16 Minuten verdoppelt werden. In dieser Zeit wurde für das zweite Experiment ungefähr die gleiche Anzahl an Grundgerüsten erzeugt, wobei allerdings jede Grundgerüstklasse deutlich mehr Repräsentanten beinhaltet, was sich durch die Verteilung der Fragmente in dem Fragmentraum erklären lässt. In dem targetspezifischen Fragmentraum war durch die geringe Fragmentanzahl die Zahl der Moleküle pro potentiellem Grundgerüst beschränkt, wohingegen es im BRICS_4k+ Fragmentraum es eine sehr hohe Anzahl Fragmente gibt, die bis auf ihre Dekoration identisch sind. (Dies sind vor allen Dingen verschieden dekorierte aromatische Heterozyklen.) Dadurch existiert für jede Grundgerüstklasse eine große Menge an möglichen Molekülen, die jeweils minimal verschiedene Dekorationen aufweisen. Solange sie den Pharmakophor erfüllen, sind diese Moleküle für Qsearch annähernd gleichwertig, was zur Folge hat, dass Änderungen, die zu äquivalenten Grundgerüsten führen, mit einer hohen Wahrscheinlichkeit akzeptiert und mit zu den Lösungen hinzugefügt werden.[3]

6.3. BCR-ABL Tyrosinkinase Inhibitoren

Chronische myelonische Leukämie und Imatinib

Die chronische myelonische Leukämie ist eine Krankheit, die durch den Austausch (die sogenannte *Translokation*) von Genabschnitten zwischen den Chromosomen 9 und 22 in pluripotenten Stammzellen entsteht.[103] Dabei wird das Chromosom 22 um einen Abschnitt verkürzt, der am Chromosom 9 wieder angefügt wird, wobei die Gene BCR und ABL1 zu einem neuen (BCR-ABL) verknüpft werden, das aufgrund seines Entdeckungsort auch *Philadelphiachromosom* genannt wird. Das durch die Fusion neu entstandene Gen wird in eine Tyrosinkinase mit erhöhter Aktivität umgesetzt. Diese Aktivität äußert sich in einer unverhältnismäßig starken Vermehrung von weißen Blutkörperchen und führt zum Krankheitsbild der chronisch myelonischen Leukämie. Ende der neunziger Jahre wurde der Wirkstoff Imatinib (Gleevec®) auf den Markt gebracht, der die krankheitsverursachenden Tyrosinkinasen gezielt blockiert und der Krebsforschung auf diesem Feld einen großen Schub gegeben hat. Das Blockieren von Tyrosinkinasen findet inzwischen in der Therapie von verschiedenen Krebsarten Anwendung.[104]

[3]Eine automatische Zusammenfassung von äquivalenten Grundgerüsten wäre an dieser Stelle hilfreich bei der Interpretation der Ergebnisse hilfreich und wird im Ausblick in Abschnitt 7.2 diskutiert.

Imatinib Pharmakophor

Für die Berechnungen wurde ein Pharmakophor für die BCR-ABL Tyrosinkinase[85] verwendet, der spezifisch für das Interaktionsmuster von Imatinib ist. In der dazugehörigen Publikation wurde gezeigt, dass mithilfe des Pharmakophors alle Imatinib Moleküle in der Protein Data Bank[90] (PDB) automatisch erkannt werden können.

Der Pharmakophor wurde auf Basis der Proteinstruktur der c-Abl Tyrosinkinase[4] im Komplex mit Imatinib (PDB:1opj[105]) reproduziert. Er besteht aus vier aromatischen Merkmalen und zwei gerichteten Wasserstoffbrückenakzeptorfunktionen. Die Platzierung der Pharmakophormerkmale auf den einzelnen Molekülteilen ist in Abbildung 6.3.1 dargestellt. Die Form wurde durch das Platzieren von Kugeln mit dem Radius 1,8Å auf allen Schweratomen definiert.

Durchführung und Ergebnisse

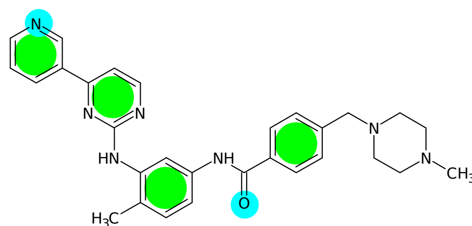
Der Pharmakophor diente zwei verschiedenen Experimenten als Grundlage. Zum einen wurde er in einer freie Suche im BRICS_4k+ Fragmentraum eingesetzt und zum anderen in einer leicht abgewandelten Form in einer Fragmenterweiterung mit einem targetspezifischen Fragmentraum.

Im ersten Anwendungsbeispiel wurde die Fähigkeit von Qsearch untersucht, chemische Modifikationen an Imatinib zu reproduzieren, die bekannterweise zu neuen aktiven Molekülen geführt haben. Dafür wurden drei Moleküle, die als Weiterentwicklungen von Imatinib auf den Markt kamen und ebenfalls Tyrosinkinasen als Zielstruktur haben, auf ihre jeweiligen Unterschiede zu Imatinib untersucht, und mit den Ergebnissen, die Qsearch geliefert hat, verglichen. Für die Berechnungen wurde Qsearch verteilt auf vier Computern je acht Stunden lang laufen gelassen, was insgesamt zu 216 Ergebnisse geführt hat. Die Ergebnisse wurden manuell ihren ähnlichsten aktiven Molekülen zugeordnet.

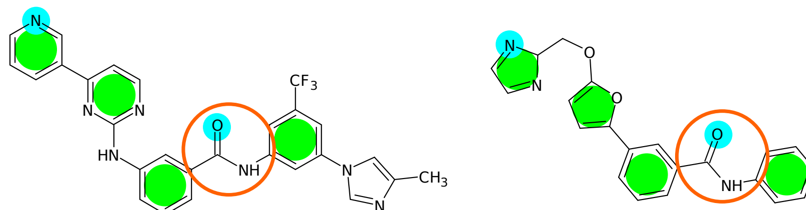
Im zweiten Anwendungsbeispiel wurde der Fragmenterweiterungsmodus von Qsearch verwendet, in dem von einem Startfragment aus ein komplettes Molekül generiert wird. Ziel war, den Grundgerüstwechsel zu reproduzieren, der in dem Imatinib-Nachfolger Ponatinib resultierte. Das Anwendungsbeispiel basiert auf einer Analyse des Interaktionsmusters das Imatinib mit dem Rezeptor in der PDB Struktur 1opj ausbildet. Die Analyse zeigt, dass keines der beiden Stickstoffatome des Pyrimidinrings Wasserstoffbrückenbindungen ausbildet, und der Ring daher in der Molekülstruktur eliminiert werden kann. Folglich wurde aus dem Pharmakophor das aromatische Merkmal entfernt und auf den

[4]Die c-Abl Tyrosinkinase ist die Tyrosinkinase, die aus dem ABL1 Gen transkribiert wird. Sie ist eine strukturelle Untereinheit der BCR-ABL Tyrosinkinase und kann daher als Zielstruktur verwendet werden.[105]

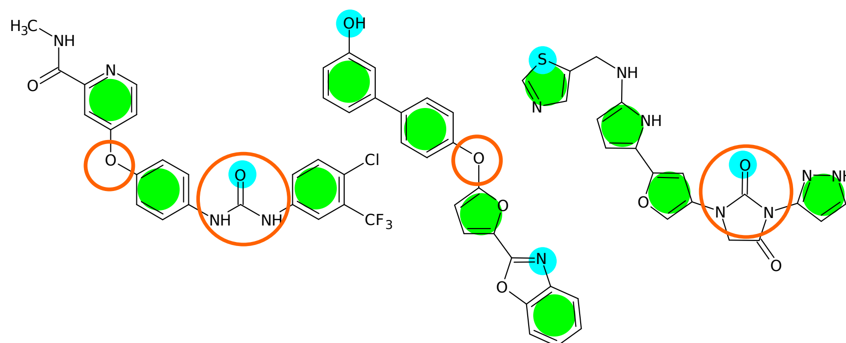
Imatinib



Nilotinib



Sorafenib



Ponatinib

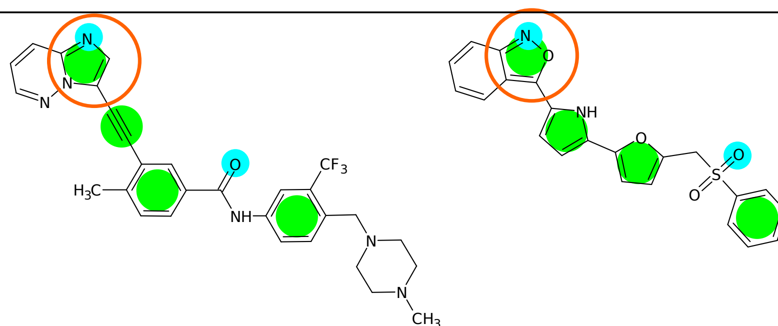


Abbildung 6.3.1.: Pharmakophor für BCR-ABL Tyrosinkinaseninhibitoren nach Wolber und Langer[85]. In der obersten Zeile ist Imatinib zusammen mit dem Pharmakophor zu sehen. Darunter abgebildet sind Wirkstoffe der zweiten Generation, die aus Imatinib hervorgegangen sind. Wichtige Unterschiede sind orange markiert. Alle vier Moleküle sind strukturell ähnlich und erfüllen den Pharmakophor entweder ganz, oder im Fall Sorafenib und Ponatinib nur teilweise. Neben den Molekülen sind jeweils Qsearch Ergebnisse abgebildet, die die gewünschten Strukturänderungen beinhalten.

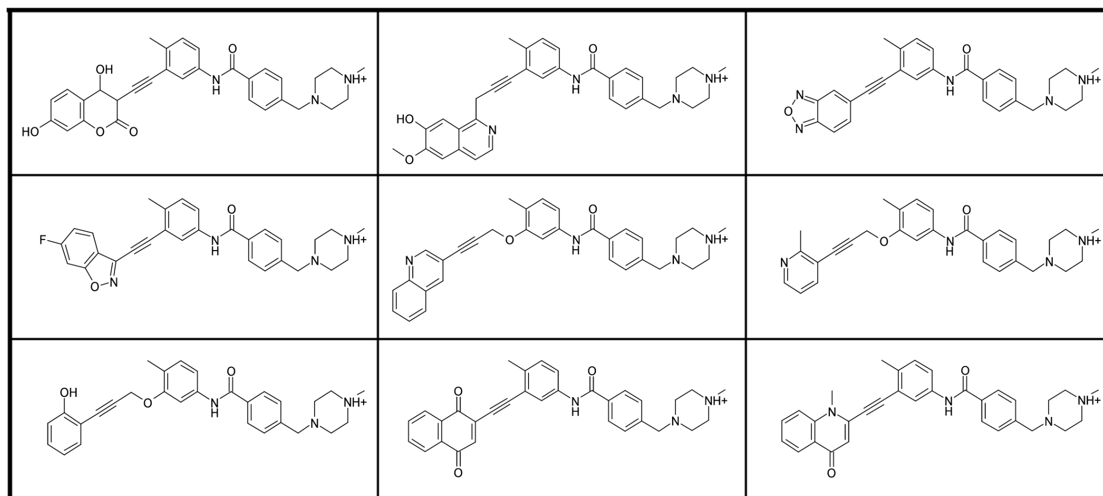


Tabelle 6.3.: Die besten neun Lösungen des Fragmenterweiterungsexperiments. Alle Lösungen wurden in das aktive Zentrum der PDB Struktur 1opj gelegt und von MOE bewertet.

Akzeptor und das aromatische Merkmal des terminalen Pyridins reduziert. Die anderen Pharmakophormerkmale sind Teil des Startfragments und wurden deshalb ebenfalls entfernt.

Als Eingabe wurde für dieses Experiment ein Unterraum des BRICS_4k Datensatzes generiert. Fragmente wurden in ihm aufgenommen, wenn sie entweder zwei oder mehr Linkatome besitzen und nicht aromatisch sind (541 von 1086 Linker-Fragmenten), oder wenn sie ein Linkatom und mindestens ein Akzeptoratom besitzen (717 von 2566 terminalen Fragmenten). In dem Experiment wurde Qsearch für neun Stunden[5] laufen gelassen, was in 2152 Ergebnissen resultierte. Nach dem Lauf wurden die Lösungen in das aktive Zentrum der Kristallstruktur platziert und mit dem Programm MOE[6] bewertet. Die neun am höchsten bewerteten Lösungen sind in Tabelle 6.3 gezeigt.

Interpretation

Zuerst werden die Ergebnisse des ersten Experiments besprochen. Sie sind nach den Molekülen geordnet, deren Grundgerüstwechsel reproduziert worden sind, und in Abbildung 6.3.1 gezeigt werden. Darauf folgt eine Analyse der Moleküle, die gänzlich neuartige Topologien aufweisen. Zuletzt werden die Ergebnisse aus der Fragmenterweiterung untersucht.

[5]Intel Xeon mit 4 Kernen zu 2,5 GHz und 4GB RAM, CentOS 5.4.
[6]MOE; Chemical Computing Group: Montreal (Kanada). `www.chemcomp.com`

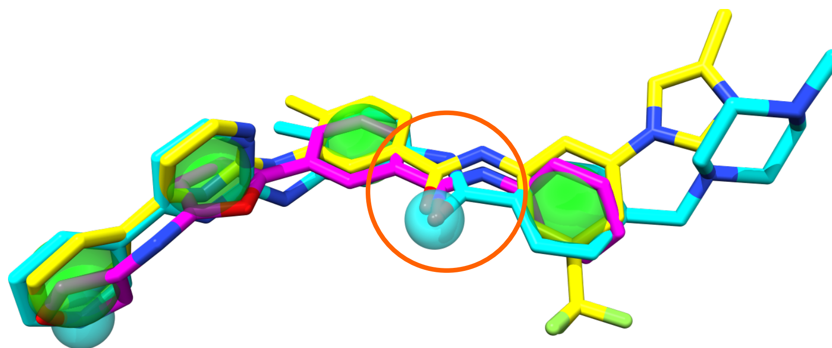


Abbildung 6.3.2.: Dreidimensionale Darstellung der bioisosteren Invertierung von Amiden (orange hervorgehoben). Die Kugeln stellen Pharmakophormerkmale dar. Die unterschiedliche Moleküle sind durch verschiedene Farben gekennzeichnet. Cyan: Imatinib, gelb: Nilotinib, magenta: Qsearch Ergebnis. Sowohl in der normalen als auch in der invertierten Form liegt das Sauerstoffatom so, dass es an der gleichen Position als Akzeptor fungieren kann.

Nilotinib Nilotinib unterscheidet sich von Imatinib hauptsächlich durch die Invertierung des zentralen Amids. Diese Invertierung ist in der medizinischen Chemie eine etablierte Technik zur bioisosteren Ersetzung von Amiden[106], da das Sauerstoffatom der Amidgruppe in beiden Orientierungen als Akzeptor fungieren kann.[7] In dem Experiment wurde genau diese Invertierung des Amids als mögliche bioisostere Lösung erzeugt.

Die Invertierung des Amids ist als Regel in Qsearch nicht explizit als bioisostere Ersetzung formuliert. Dass sie trotzdem reproduziert werden konnte, zeigt das Potential der Suche mit dreidimensionalen Pharmakophoren in chemischen Fragmenträumen, um bioisostere Ersetzungen ohne *a priori* Wissen zu erzeugen. So können durch die Methode, auch nicht offensichtliche bioisostere Ersetzungen generiert werden, was eine wichtige Voraussetzung für Grundgerüstwechsel darstellt.

Sorafenib Sorafenib weist im Vergleich zu Imatinib eine Reihe an Modifikationen auf. Mit dem Pharmakophor kompatibel sind in diesem Beispiel nur die Ersetzung des Amins durch einen Ether und der Austausch des Amids durch eine Harnstoffgruppe.

Erstere Ersetzung konnte ohne Probleme reproduziert werden und findet sich als häufiges Muster in den generierten Lösungen. Allerdings enthält der Amin/Ether Austausch keine relevante Information, wenn das Hauptinteresse auf dem Generieren von neuartigen Grundgerüsten liegt, da beide Gruppen ähnliche geometrische und topologische

[7] Auch die Donorfunktion des Stickstoffs bleibt in beiden Orientierungen erhalten, was in dem hier besprochenen Fall allerdings nicht relevant ist.

Eigenschaften aufweisen.

Die interessantere Ersetzung des Amids durch eine Harnstoff-Gruppe hingegen wurde von Qsearch nicht vorgeschlagen. Stattdessen findet sich an dieser Stelle häufig ein zyklische Hydantoin-Ring, der dem Harnstoff sehr ähnlich ist. Diese Beobachtung lässt sich auf den verwendeten Fragmentraum zurückführen. Die Schnittregeln, die eingesetzt wurden, um ihn zu erstellen, trennen sämtliche Harnstoffgruppen an beiden Amidbindungen auf. Dadurch entsteht ein Carbonyl-Fragment mit zwei Linkatomen, sowie zwei weitere Fragmente, die jeweils an den Stickstoffatomen einen Linker besitzen. Da jedes Fragment in den BRICS Fragmenträumen nur genau einmal vorkommen darf, existiert auch das Fragment, das die Carbonylgruppe bereitstellt, die zur Bildung einer Harnstoffgruppe benötigt wird, nur ein Mal. Zyklische Harnstoffe hingegen werden nach den BRICS-Regeln nicht geschnitten, weswegen das Hydantoin in verschiedenen chemischen Umgebungen im Fragmentraum weitaus öfter vorkommt (in diesem Fall 30 Mal). Die Platzierung des Carbonyl Fragments an der richtigen Stelle und die Kombination mit den korrekten Partner, ist im Vergleich zur Platzierung eines kompletten zyklischen Harnstoffs relativ unwahrscheinlich, und erklärt somit die gefundenen Ergebnisse.

Ponatinib Die Struktur von Ponatinib ist chemisch weitaus entfernter von Imatinib die der vorherigen Beispiele. Trotzdem teilt es ihre grundsätzliche Linearität im topologischen Aufbau des Grundgerüsts. Die in Ponatinib eingeführten Änderungen beinhalten unter anderem die Ersetzung des Pyrimidinrings durch Ethinyl und die des terminalen Pyridins durch einen anellierten aromatischen Heterozyklus. Keine der Lösungen beinhaltet die Ethinylersetzung, da an dieser Stelle vom Pharmakophors ein aromatischer Ring gefordert wird. Allerdings wurde der terminale Pyridinring durch ein Benzooxazol ersetzt, das sehr ähnlich zu dem Heterozyklus ist, der in Ponatinib zu finden ist.

Neuartige und nicht-lineare Grundgerüste Zusätzlich zu den Variationen in den Topologien der Grundgerüste, die die bekannten aktiven Moleküle aufweisen, und die allesamt linear sind (vgl. Abbildung 6.3.1), wurden neuartige, nicht-lineare, Grundgerüste generiert. Zwei Beispiele sind in Abbildung 6.3.3 dargestellt. Das erste Molekül führt mithilfe eines sp^3-hybridisierten Kohlenstoffatoms eine neue Verzweigung ein, und das zweite Molekül erhält durch ein bizyklisches System mit ortho-Substitution eine neuartige Topologie.

Werden diese Lösungen in das aktive Zentrum der Zielstruktur platziert und analysiert, so bilden diese Moleküle nicht alle Interaktionen aus, die bei den bekannten Wirkstoffen wie zum Beispiel bei Imatinib beobachtet werden können. Dies ist ein erwartbares Ver-

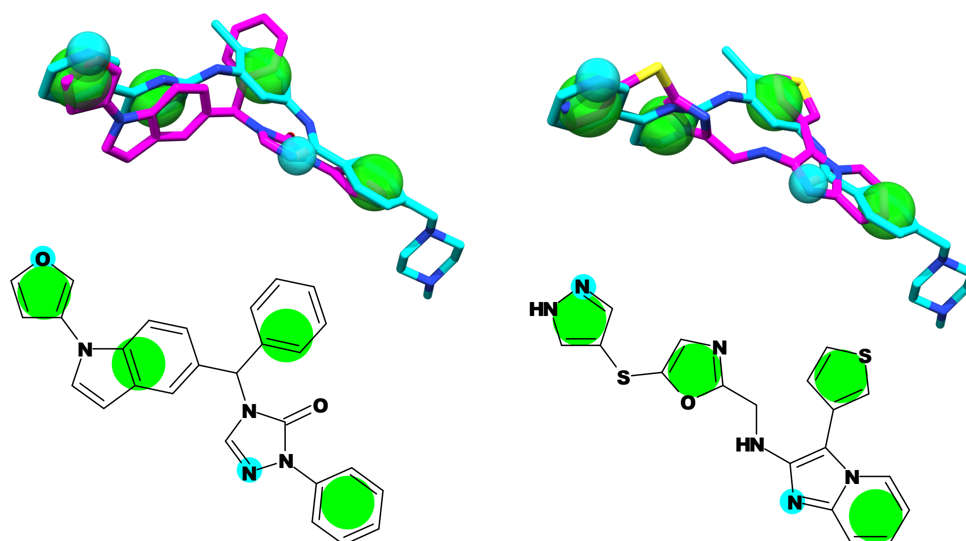


Abbildung 6.3.3.: Generierte Lösungen für die BCR-ABL Tyrosinkinase mit neuartigen Grundgerüsten. Die Moleküle sind unten mit ihrer Strukturformel angegeben und oben in magenta in 3D dargestellt. In der 3D Abbildung ist jeweils Imatinib in cyan zu sehen. Grüne Kugeln repräsentieren aromatische Pharmakophormerkmale, cyanfarbene Kugeln Akzeptormerkmale.

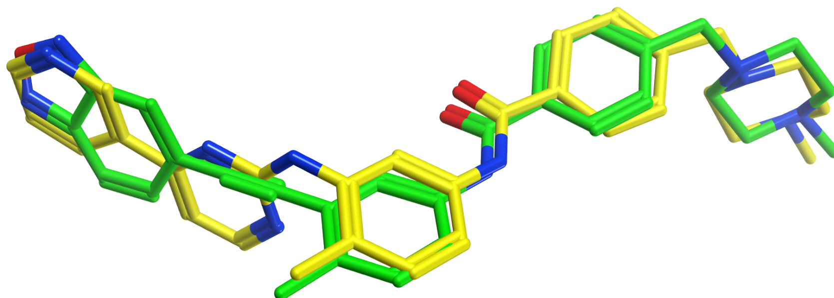


Abbildung 6.3.4.: In der Fragmenterweiterung generiertes Molekül (Molekül auf Rang 5, grün) überlagert mit Imatinib (gelb). Das meta-substituierte Pyrimidin wurde durch ein Ethinyl, und das terminale Pyridin durch ein Benzo[1,2,5]oxadiazol unter Erhaltung des Wasserstoffbrückenbindungsmusters ersetzt. Es ist zu erkennen, dass es dem Startfragment erlaubt war, sich leicht zu bewegen.

halten, da sie nicht Teil der Pharmakophorhypothese waren. Um die Interaktionen mit der Zielstruktur zu optimieren, müssten die gefundenen Lösungen entweder manuell oder mit anderen computergestützten strukturbasierten Methoden verbessert werden.

Am Beispiel der Lösung mit dem sp^3-hybridisierten Kohlenstoffatom seien hier mehrere potentielle Optionen für eine Optimierung dargelegt: Bei der Analyse der dreidimensionalen Struktur von Imatinib in der Kristallstruktur (PDB:1opj) ist ersichtlich, dass das zentrale Amid eine Wasserstoffbrückenbindung mit dem N-H des Proteinrückgrats ausbildet. Folglich könnte das Triazolon des Grundgerüsts durch ein Imiadzol ersetzt werden, das die Donor- und Akzeptoratome in der korrekten Position für das Interaktionsmuster hat. Darüber hinaus könnte auf Basis der Beobachtung, dass sowohl der Pyridin-Stickstoff von Imatinib als auch sein benachbarter C-H azide Kohlenstoff ebenfalls Wasserstoffbrücken ausbilden, der terminale Furanring des generierten Moleküls durch einen 3-substituierten Pyrazolring ersetzt werden, um die Interaktionen zu optimieren.

Fragmenterweiterung Die Lösungen des zweiten Experiments wurden dadurch analysiert, dass sie in das aktive Zentrum der Tyrosinkinase platziert wurden und mit dem Programm MOE auf Basis der ausgebildeten Interaktionen bewertet wurden. Die neun besten Lösungen sind in Tabelle 6.3 zu sehen. Alle Lösungen enthalten ein Ethinyl oder ein Ethinyl mit leichten Modifikationen als Verbindungsstück, welches genau die Ersetzung ist, die zu Ponatinib geführt hat. Aus diesen Ergebnissen ist ersichtlich, dass mithilfe von Qsearch die acht Fragmente mit einem Ethinylmotiv aus den insgesamt 541 Linker-Fragmenten als die geometrisch besten identifiziert werden. Die Kristallstruktur von Imatinib ist in Abbildung 6.3.4 mit der fünftplatzierten Lösung gezeigt. Das gene-

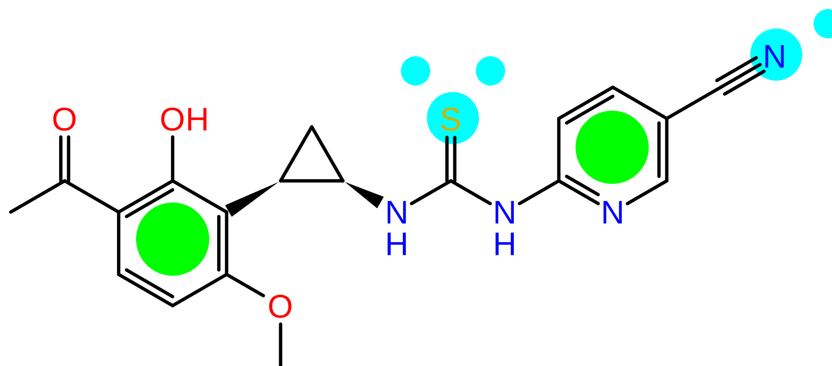


Abbildung 6.4.1.: Der Reverse-Transkriptase-Inhibitor MSC194 (PDB:1iky[107]) mit dem für die Parametrisierung verwendeten Pharmakophor. Auf dem Nitril ist ein Akzeptormerkmal (cyan) mit einer Richtung und auf dem Thioharnstoff eines mit zwei Richtungen platziert. Die Richtungen der aromatischen Merkmale (grün) sind nicht dargestellt.

rierte Molekül zeigt einen sehr guten Volumenüberlapp, und entspricht allen Vorgaben des Pharmakophors.

6.4. Parametrisierung von Qsearch

Um die Simulated Annealing Prozedur in Qsearch zu optimieren wurden diverse Parametersätze erprobt. Allerdings konnte, die Auswahl nicht durch eine fundierte statistische Analyse durchgeführt werden. Grundlage hierfür ist eine hohe Anzahl an Messpunkten, die nur durch eine automatische Bewertung der Ergebnisse möglich ist.

Da die Qualität der Ergebnisse nicht automatisch bewertbar ist, wurde die Anzahl der generierten Moleküle als Optimierungskriterium verwendet. Im folgenden werden die Experimente beschrieben, die für die Parametrisierung verwendet wurden.

Der Pharmakophor für die Studien wurde mithilfe der Struktur des Reverse-Transkriptase-Inhibitors[8] MSC192 (PDB:1iky[107]) erstellt. Er ist in Abbildung 6.4.1 skizziert und besteht aus zwei gerichteten Akzeptormerkmalen sowie zwei gerichteten aromatischen Merkmalen. Die Form wurde durch das Platzieren von Kugeln mit einem Radius von 1,8Å auf allen Schweratomen definiert.

Als Qualitätskriterium für die Optimierung wurde die Anzahl der generierten Lösungen

[8]Die reverse Transkriptase ist für die Umwandlung von einsträngiger RNA in einsträngige DNA verantwortlich. Sie wird von Retroviren dazu verwendet, ihr RNA-Genom in DNA zu transkribieren, woraufhin es über Integrasen in das Genom ihrer Wirtszellen eingebaut werden kann. Die Hemmung der reversen Transkriptase kommt unter anderem in der Behandlung von HIV im Rahmen der hochaktiven antiretroviralen Therapie zum Einsatz.[108]

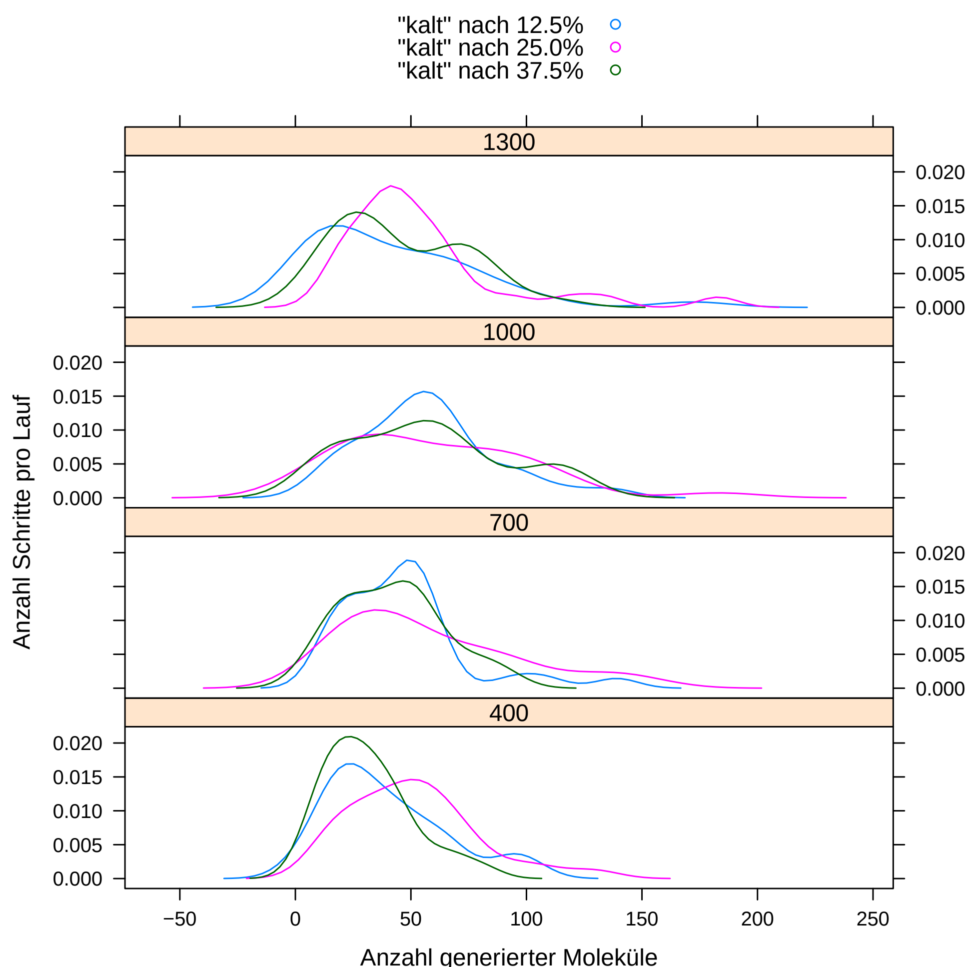


Abbildung 6.4.2.: Kerndichteschätzung für die Anzahl der generierten Lösungen aufgeteilt nach den Parameterkombinationen.

nach fünf Minuten gewählt. Die variierten Parameter beinhalten die Anzahl der Schritte pro Lauf (400, 700, 1000, 1300) und den Abkühlungsplan („kalt“ nach 12,5%, 25,0%, 37,5% der Schritte), wobei „kalt“ in dem Algorithmus einen Zustand beschreibt, der eine Temperatur nahe Null hat und in dem nur noch Lösungen akzeptiert werden, die zu einer Verbesserung führen. Als Eingabe diente der oben beschriebene Pharmakophor und der BRICS_4k+ Raum.

Für jede Kombination an Parametern wurden 30 Läufe zu je fünf Minuten durchgeführt und die Anzahl der gefundenen Lösungen mithilfe einer Kerndichteschätzung in Abbildung 6.4.2 analysiert. Weder in der Abbildung noch in den Zahlen selbst ließ sich ein offensichtlicher Trend erkennen.

Bei der visuellen Inspektion der generierten Moleküle wurde darüber hinaus klar, dass die Anzahl der Lösungen alleine kein geeignetes Kriterium ist um die Qualität eines Parametersatzes zu beschreiben. Die Ergebnisse sind – gerade wenn eine hohe Anzahl an Schritten pro Lauf verwendet wurde – von einer großen Zahl von Molekülen mit nur minimalen Variationen dominiert. Dies ist eine Folge der langen „kalten“ Phasen, in der nur Ersetzungen mit beinahe identischen Fragmenten stattfinden, was insbesondere für Heteroaromaten beobachtet werden konnte. Ein weitaus besseres Maß wäre die Anzahl an verschiedenen Grundgerüsten, die pro Lauf generiert werden. Allerdings gibt es bisher noch keine automatische Beschreibung von Grundgerüsten, die befriedigende Ergebnisse liefert (vgl. Abschnitt 2.4), und eine visuelle Analyse der Ergebnisse wäre durch ihre hohe Anzahl sehr aufwendig und darüber hinaus stark subjektiv.

Statt eines statistisch fundierten Parametersatzes werden für Qsearch daher Werte verwendet, die aus den Erfahrungen die im Laufe des Projektes gemacht wurden, resultieren, und die durch die Beobachtung des Algorithmus bei der Optimierung geprägt sind. So sind kürzere Läufe mit schnellen Abkühlungszeitplänen zu bevorzugen. Längere Läufe resultieren darin, dass viele Variationen eines Grundgerüsts generiert werden, und langsamere Abkühlungen darin, dass insgesamt weniger Grundgerüste generiert werden, da der Algorithmus mehr Zeit in einem Stadium verbringt, in dem praktisch alle Änderungen angenommen werden. Daher werden die einzelnen Optimierungsläufe mit jeweils 700 Schritten durchgeführt und die Parameter so gewählt, dass die Temperatur nach 12,5% „kalt“ ist. Für die Starttemperatur wurde der Wert 8,0 gewählt.

6.5. Einschränkungen der Methode

6.5.1. Hydrophobizität

Eine wichtige Einschränkung des verwendeten Chemiemodells ist die Abbildung der Hydrophobizität. Sie wird durch die IUPAC folgendermaßen definiert:

> Als Hydrophobizität wird das Aneinanderlagern von nicht polaren Gruppen oder Molekülen in wässriger Umgebung bezeichnet, die durch die Neigung von Wasser, nicht-polare Moleküle auszuschließen, verursacht wird. (übersetzt aus [109])

Hydrophobe Interaktionen beruhen auf zwei Effekten, wobei die Stärke des ersteren im Vergleich zum zweiten vernachlässigbar ist. Zum einen induzieren hydrophobe Regionen sich gegenseitig Bipole, die sich dann über ihre komplementären Ladungen anziehen. Dieser Effekt wird unter dem Namen „London-Kräfte“ geführt und ist unter anderem

für den Zusammenhalt von apolaren Flüssigkeiten verantwortlich. Zum anderen werden hydrophobe Regionen in polaren Lösungsmitteln wie Wasser ausgeschlossen, was zum Beispiel bei den Phasentrennungen von Öl-Wasser Gemischen zu beobachten ist.[91]

In der Wirkstoffforschung spielt die Hydrophobizität in zwei entgegenstrebenden physikochemischen Eigenschaften eine wichtige Rolle. Einerseits hat sie Einfluss auf die Wasserlöslichkeit der Wirkstoffe, wobei sie sich umgekehrt proportional zur Hydrophobizität verhält, und eine wichtige Grundlage für die Möglichkeit zur oralen Darreichung darstellt. Andererseits beeinflusst die Hydrophobizität die Bindungsaffinität eines Wirkstoffs, wobei sie in der Regel proportional mit ihr steigt. Eine weit verbreitete Praxis um die Affinität von Wirkstoffen zu erhöhen, ist daher das Einführen von lipophilen Gruppen in ein Molekül.[97] Das Austarieren der Eigenschaften von Wirkstoffen sowohl in Bezug auf eine hohe Wasserlöslichkeit als auch auf eine hohe Affinität ist eine der Kernaufgaben der medizinischen Chemie.[110]

Hydrophobe Regionen werden komplementär zu der Zielstruktur postuliert und finden sich in vielen publizierten Pharmakophoren wieder. (z. B. für ACE-2[12], COX-2[54], oder Histamin H-3 Rezeptoren[111]) Für den in den Ergebnissen präsentierten COX-2 Pharmakophor existieren verschiedene Variationen, in denen die aromatischen Regionen teilweise oder ganz durch hydrophobe Pharmakophormerkmale ersetzt sind.[9]

Qsearch beschränkt sich darauf, das Ausschließen von hydrophoben Regionen in wässriger Lösung zu modellieren, da die London-Kräfte im Verhältnis sehr schwach sind. Der Effekt wird durch die Abwesenheit von hydrophilen Interaktionen modelliert. Das heißt, ein hydrophobes Merkmal ist dann erfüllt, wenn mindestens ein Atom ohne und kein Atom mit hydrophilen Interaktionen in ihm liegt.

Das in Qsearch verwendete Konzept unterscheidet sich wesentlich von dem vergleichbarer Programme, die versuchen, die Idee von „hydrophoben Gruppen“ zu verwirklichen, und die hydrophobe Punkte auf Basis von funktionellen Gruppen oder entlang von Kohlenstoffketten platzieren. Ein direkter Vergleich der Methoden findet sich in Abbildung 6.5.1. Unter den Anwendern dieses Ansatzes finden sich unter anderem die kommerziellen Programme Catalyst[73], Phase[10], LigandScout[85] und MOE.[112] Der Ansatz der hydrophoben Gruppen bietet zwei wichtige Vorteile. Zum einen lässt sich so die Hydrophobizität auf einen einzelnen Punkt lokalisieren, was für viele Algorithmen Vorteile bietet, und zum anderen wird das Konzept der hydrophoben Gruppen in der medizinischen Chemie häufig verwendet, weswegen die Programme das Konzept in ihren Modellen

[9] Aromatische Ringe werden in der Literatur oft als hydrophob interpretiert, auch wenn sie polare Interaktionen ausbilden können.[112]

[10] Phase; Schrödinger, LLC: New York, NY (USA). `www.schrodinger.com`

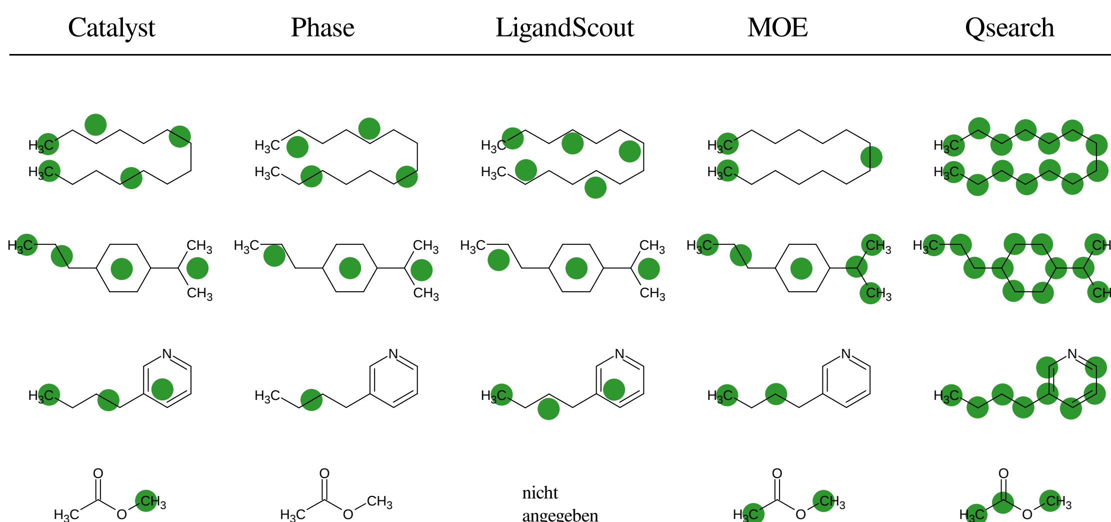


Abbildung 6.5.1.: Nebeneinanderstellung verschiedener Hydrophobizitätskonzepte nach Spitzer et al.[18] und Wolber et al.[112]. Für jedes Programm sind in den Spalten anhand verschiedener Moleküle die Platzierungen der lokalisierten Beschreibung der Hydrophobizität (dunkelgrün) dargestellt. Qsearch verwendet als einzige Methode eine atombasierte Beschreibung, in der die hydrophoben Eigenschaften nicht zusammengefasst werden. Die anderen Methoden platzieren die hydrophoben Eigenschaften als Kugeln nach funktionellen Gruppen und/oder entlang von Kohlenstoffketten.

abbilden müssen.

Für Qsearch wurde bewusst ein anderer Ansatz gewählt um die Hydrophobizität zu beschreiben. Dafür ausschlaggebend waren einerseits die vielen unbeantworteten Fragen beim Konzept der hydrophoben Gruppen[11] und andererseits die schlechte Eignung des Konzepts für das *de novo* Design (mit dem sich die o.g. Programme nicht beschäftigen). Beim *de novo* Design lassen sich nämlich in der Regel polare Strukturmotive, in einem Nachbearbeitungsschritt einfach durch geometrisch äquivalente apolare Strukturen ersetzen, wenn sie in Regionen liegen, in denen hydrophobe Elemente erwünscht sind – im einfachsten Fall dadurch, dass die Heteroatome durch Kohlenstoff ersetzt werden. Die zu frühe Berücksichtigung von hydrophoben Regionen hingegen schränkt die Freiheit des Algorithmus bei der Generierung von neuen Grundgerüsten stark ein, insbesondere, wenn die verwendeten Fragmente (wie die der BRICS-Räume) sehr polar sind.

Trotzdem ist eine Verbesserung der Behandlung von hydrophoben Merkmalen wichtig. Auch wenn sie keinen Vorteil bei der Generierung von neuen Grundgerüsten bieten, so ist doch die Akzeptanz der Lösungen bei den Benutzern höher, wenn sich in den von ihnen postulierten hydrophoben Regionen entsprechende Gruppen finden. Dies könnte zum Beispiel durch einen der Suche nachgelagerten Schritt automatisiert werden, in dem die Lösungen den Erwartungen angepasst werden.

6.5.2. Torsionsmodell

Ein wichtiges Kriterium für die Qualität der Ergebnisse ist, dass die Konformationen der von Qsearch erstellten Moleküle physikalischen Gegebenheiten entsprechen. Studien, in denen die hinterlegten Strukturen der Cambridge Structural Database[116] statistisch ausgewertet wurden, zeigen, dass es entlang bestimmter chemischer Bindungen präferierte Torsionswinkel gibt, die in Molekülen eingenommen werden.[117, 118] In Qsearch wird hingegen angenommen, dass alle Winkel gleichberechtigt sind. So können in den generierten Molekülen ungünstige Konformationen vorkommen, die es dem Molekül zwar erlauben, den Pharmakophor zu erfüllen, aber durch ihre hohe interne Energie unwahrscheinlich sind. Da Wirkstoffe in der Regel in niederenergetischen Konformationen an ihre Zielstrukturen binden[119], ist es jedoch wünschenswert, dass die generierten Strukturen ebenfalls nahe an energetischen Minima liegen und keine ungünstigen Konformationen

[11] Das punktbasierte Konzept lässt viele Fragen offen.[18] Unter anderem: Wie werden komplexe hydrophobe Geometrien beschrieben? Wie werden die Punkte am besten in ausgedehnten hydrophoben Regionen platziert? Und welche Atome sollten als hydrophob angesehen werden? Antworten auf diese Frage werden unter anderem in der Strukturvorhersage von Proteinen gesucht, wo der hydrophobe Effekt als eine der treibenden Kräfte entdeckt wurde[113], und teilweise auch im Scoring von Protein-Ligand Komplexen, wo seine Betrachtung ebenfalls von hoher Relevanz ist.[114, 115]

aufweisen.

Die zu Qsearch vergleichbaren Ansätze beinhalten allesamt einen Term, um die Wahrscheinlichkeit von bestimmten Torsionswinkeln zu beschreiben und darauf basierend Torsionswinkel zu wählen. Allerdings ist diese Berechnung in der Regel aufwendig, und kann die Laufzeit der Programme negativ beeinträchtigen. SkelGen[30] verwendet den Torsionsterm des COSMIC Kraftfeldes[120], hat ihn allerdings aus Effizienzgründen standardmäßig ausgestellt, BIBuilder[38] benutzt das in MOE verwendete Kraftfeld, um die generierten Strukturen zu minimieren, was aber erst im letzten Schritt des Algorithmus passiert, und NovoFLAP[65] umgeht das Problem dadurch, dass es das externe Programm OMEGA verwendet, um Konformationen mit regelbasiert zugewiesenen Torsionswinkeln zu erstellen.

Eine weitere Vereinfachung der Handhabung von Konformationen stellt in Qsearch die vereinfachte Modellierung der Rotierbarkeit an Bindungen dar. Doppelbindungen werden von Qsearch als fix angesehen, auch wenn an ihnen zwei mögliche Konformationen möglich sind. So kann eine Reihe an Konformationen durch den Ausrichtungsalgorithmus nicht erreicht werden.

Eine aufwendigere Behandlung des Torsionsterms in Qsearch würde die Qualität der Ergebnisse dadurch verbessern, dass die erzeugten Ergebnisse realistischere Konformationen besitzen würden. Dafür ist es denkbar, entweder den Term eines Kraftfeldes in der Molekülausrichtungsphase mit heranzuziehen, oder auf Basis statistischer Analysen eine Pseudo-Energie für Torsionswinkel zu berechnen, die ihre relative Häufigkeit beschreibt. Hierfür muss jedoch eine Abwägung zwischen dem Rechenaufwand und den erwarteten Qualitätssteigerungen getroffen werden.

6.6. Fragmenträume und ihre Entdekorierung

Der in den Methoden vorgestellte Entdekorierungsschritt hilft Qsearch auch mit stark dekorierten Fragmenträumen umzugehen. Allerdings ist dieser Mechanismus nicht für alle Fragmenträume gleichermaßen geeignet. Um die Eignung dieser Methode in Bezug auf die verschiedenen Arten von Fragmenträumen zu beurteilen, werden im folgenden ihre speziellen Eigenschaften und Anwendungsfelder vorgestellt und dann als Bewertungsgrundlage verwendet. Eine Zusammenfassung der Eigenschaften findet sich in Tabelle 6.4.

kombinatorische Fragmenträume sind den konkreten Syntheseprotokollen in kombinatorischen Bibliotheken nachempfunden.[45] Die Fragmente korrespondieren zu realen Reagenzien, wie sie im Labor für die Synthese verwendet werden würden. Mo-

Typ	Ursprung	implizites Wissen	Dekoration	Anwendung	Entdekorierung
kombinator. Raum	Reagenzien	Synthesen	reaktionsspezifisch	konkrete Synthese	nicht geeignet
targetspez. Raum	aktive Moleküle	Interaktionsmuster	targetspezifisch	Rekombination	teilw. geeignet
generischer Raum	Wirkstoffbibliotheken	Wirkstoffähnlichkeit	unspezifisch	Ideengenerator	geeignet

Tabelle 6.4.: Drei verschiedene Arten von Fragmenträumen und ihre Eigenschaften.

leküle, die aus ihnen generiert werden, können in der Regel direkt synthetisiert werden.

targetspezifische Fragmenträume werden aus einer kleinen Menge an Molekülen gewonnen für die bekannt ist, dass sie Aktivität gegenüber einer Zielstruktur aufweisen. Die Fragmente werden durch retrosynthetische Zerlegung erzeugt und weisen in der Regel molekulare Muster auf, die eine Aktivität begünstigen. Eine Rekombination dieser Muster kann zu neuen Molekülen mit verbesserten Eigenschaften führen. Targetspezifische Fragmenträume wurden in verschiedenen Projekten erfolgreich dazu verwendet, neuartige aktive Moleküle zu generieren. Publikationen für die erfolgreiche Anwendungen finden sich unter anderem für die Programme E-novo[121], BREED[39] und ADAPT[122].

generische Fragmenträume werden durch die retrosynthetische Zerlegung von Wirkstoffbibliotheken gewonnen. Bekannte Vertreter sind unter anderem der COBRA- und der BRICS-Raum. [43, 32] Sie zielen darauf ab, möglichst diverse molekulare Muster und Topologien bereit zu stellen, die eine hohe Ähnlichkeit zu Wirkstoffen haben. Moleküle, die aus ihnen entstehen, müssen in der Regel vor der Synthese durch Chemiker manuell modifiziert werden. (vgl. [123, 64])

In den Fragmenträumen, die Syntheseprotokolle der kombinatorischen Chemie modellieren, sollte die Entdekorierung nicht eingesetzt werden. Eine Veränderung der Reagenzien hätte Auswirkungen auf die Durchführung der Synthesen, die in diesem Raum implizit modelliert worden sind. Da der größte Vorteil dieser Räume die einfache Zugänglichkeit zu den in ihnen enthaltenen Molekülen ist, ist die Entdekoration der Fragmente in diesem Fall kontraproduktiv.

Die Entdekoration von targetspezifischen Fragmenträumen ist ein Grenzfall. Die einzelnen Fragmente korrespondieren zwar nicht zu einzelnen Reagenzien einer Synthese,

aber viele der funktionellen Gruppen, die in ihnen enthalten sind, sind in der Interaktion mit der Zielstruktur involviert. Das Entfernen dieser Gruppen würde den Verlust von wichtiger Information nach sich ziehen. Abhängig von dem Ursprung der Fragmente können die funktionelle Gruppen allerdings auch hier nur für die Modulation von pharmakokinetischen Eigenschaften angefügt worden sein, so dass ihre Entfernung kein Problem darstellen würde. Durch einen sehr spezifischen Pharmakophor ließe sich dieser Effekt kompensieren, weil funktionelle Gruppen, die für seine Erfüllung (also die Interaktion mit der Zielstruktur) erforderlich wären, durch den Algorithmus nicht entfernt werden würden.

Lediglich in den generischen Fragmenträumen lässt sich der Entdekorierungsschritt ohne potentielle Komplikationen durchführen. Die Informationen, die implizit in ihnen enthalten sind, sind sehr allgemeiner Art. Die generischen Fragmenträume dienen hauptsächlich als Quelle für möglichst diverse Moleküle, die zumindest theoretisch über einfache Synthesen zu erstellen sind, da sie häufige molekulare Motive und Verknüpfungen zwischen ihnen beinhalten. Die funktionellen Gruppen und molekularen Motive, die in ihnen vorkommen, wurden teilweise für die Modulation von pharmakokinetischen Eigenschaften angefügt, beziehungsweise tragen ohne ihre jeweils zugehörige Zielstruktur keine relevanten Information mehr in sich, und können daher ohne Probleme entfernt werden.

7. Zusammenfassung und Ausblick

In dieser Arbeit wurde eine neue Methode für virtuelles *de novo* Design von Wirkstoffen vorgestellt und die breite Anwendbarkeit der Methode anhand von vier Anwendungsszenarien gezeigt. Im folgenden werden sowohl die Methodik als auch die erzielten Ergebnisse abschließend zusammengefasst und kritisch beleuchtet. In dem darauf anschließenden Ausblick sind interessante und verfolgenswerte Aspekte des *de novo* Designs zusammengetragen, die sich während der Forschungsarbeiten ergeben haben, aber außerhalb der Domäne der durchgeführten Arbeiten befinden.

7.1. Zusammenfassung

Methodik

Die unter dem Namen Qsearch vorgestellte Methode ist die einzige vollständige Beschreibung eines Ansatzes, in dem stochastische Verfahren verwendet werden, um chemische Fragmenträume mithilfe von Pharmakophoren zu durchsuchen.[1] Die herausragenden Vorteile der verwendeten Konzepte sind im folgenden aufgeführt:

Verwendung eines Pharmakophors

- Die Beschreibung der für die *molekulare Erkennung* notwendigen Eigenschaften mithilfe eines dreidimensionalen Pharmakophors ist unabhängig von der Topologie der Moleküle, die für seine Erstellung herangezogen wurden. Dies ist eine gute Grundlage für das Erzeugen neuartiger Grundgerüste ohne durch die Strukturen von bekannten Molekülen voreingenommen zu sein.

- Das Pharmakophorkonzept ist in der medizinischen Chemie weit verbreitet und stellt durch seine intuitive Formulierung eine niedrigschwellige Methode dar, die Ergebnisse zu interpretieren. Dies ermöglicht einen leichten interdisziplinären Austausch, wie er in der Wirkstoffforschung vonnöten ist.[3]

[1] Die Publikationen, in denen die Ergebnisse von SkelGen[54] und NovoFLAP[65] vorgestellt werden, erwähnen ähnliche Ansätze, gehen allerdings nicht näher auf die verwendeten Methoden ein, um ihr intellektuelles Eigentum in Form von proprietäre Algorithmen zu schützen.

- Im Vergleich zu interaktionsbasierten Straffunktionen, wie sie in der strukturbasierten Wirkstoffsuche verwendet werden, ist die Berechnung von pharmakophorbasierten Straffunktionen weniger rechenintensiv und ermöglicht das Bewerten von mehr Molekülen in gleicher Zeit.
- Die Verwendung von Pharmakophorhypothesen ist selbst dann möglich, wenn keine dreidimensionalen Koordinaten für die Zielstruktur vorliegen.

Verwendung von Fragmenträumen

- Fragmenträume bieten durch die Verwendung von retrosynthetischen Erstellungs- und Verknüpfungsregeln eine erprobte Möglichkeit, Moleküle zu generieren, die eine höhere Wahrscheinlichkeit haben, chemisch valide und leicht synthetisierbar zu sein, als atombasierten Verfahren.[27]
- Die einfache Erstellung von targetspezifischen Fragmenträumen erlaubt es in einer intuitiven Art und Weise den Suchraum auf relevante Bereiche einzuschränken.
- Der in den Methoden vorgestellte Entdekorierungsschritt erweitert die verwendeten Fragmenträume um Moleküle, die für einen konkreten Pharmakophor besonders interessant sind.

Ergebnisse

Qsearch wurde in vier Szenarien mit diversen Parametern erprobt, um verschiedene Aspekte der Methode beurteilen zu können. Dabei wurde einerseits anhand eines COX-2 Pharmakophors die Fähigkeit der Methode dargestellt, Grundgerüste von bekannten Molekülen zu generieren, und andererseits anhand eines Pharmakophors für die BCR-ABL Tyrosinkinase, die Möglichkeit bekannte Grundgerüständerungen zu reproduzieren und neuartige Grundgerüste vorzuschlagen. In allen Szenarien lieferte Qsearch Ergebnisse, die auf eine breite Anwendbarkeit der Methode schließen lassen.

7.2. Ausblick

Die in Kapitel 6 vorgestellten Ergebnisse zeigen die Anwendung von Qsearch in einem breiten Spektrum von verschiedenen *de novo* Designszenarien. Trotzdem verbleiben Einschränkungen, deren Behandlung den Rahmen der durchgeführten Arbeiten überstiegen hätten sowie interessante Aspekte, die sich während der Arbeiten aufgetan haben, und die im folgenden zusammengetragen sind.

Dynamische Auswahl von Ringkonformationen Die Wechsel der Konformationen in nicht-aromatischen Ringen stellt eine wichtige Quelle für Variabilität in den dreidimensionalen Strukturen von Wirkstoffen dar. Qsearch modelliert die Flexibilität dieser Ringsysteme durch mehrfache Einbindung der betreffenden Fragmente in verschiedenen Konformationen in den Fragmentraum. Dieser Ansatz hat den Nachteil, dass der Fragmentraum um chemisch identische Fragmente vergrößert wird. Eine Behandlung der Ringflexibilität in der Molekülausrichtungsphase würde den Fragmentraum verkleinern und den Suchraum so weniger redundant gestalten. Dabei könnte entweder eine Datenbank mit verschiedenen Ringkonformationen oder die dynamische Erstellung der Konformationen mithilfe eines Kraftfeldes Anwendung finden.

Erstellung neuer Fragmenträume In der Präsentation der Ergebnisse in Abschnitt 6.1 wird dargelegt, wie die BRICS-Räume für die Verwendung in Qsearch nicht immer optimale Eigenschaften aufweisen. Um die Ergebnisse von Qsearch weiter zu verbessern, nimmt die Erstellung von Fragmenträumen, die für das *de novo* Design optimiert worden sind, eine Schlüsselrolle ein. Folgende Eigenschaften sollten bei der Fragmentzusammenstellung besondere Bedeutung erfahren:

- Topologische Diversität durch komplexe Verzweigungen innerhalb der Fragmente und/oder mehr Linkatome in den Fragmenten
- Chemische Diversität durch eine ausgeglichenere Auswahl von Heterozyklen
- Geometrische Diversität durch die bevorzugte Auswahl von Fragmenten mit nicht-planaren Ringsystemen

Durch eine höhere Diversität bei den Eingabefragmenten können die von Qsearch generierten Moleküle ein größeres Feld des chemischen Raums abdecken, womit die Ergebnisse substantiell verbessert werden könnten.

Grundgerüstgruppierung Je nach eingesetztem Pharmakophor und Fragmentraum kann Qsearch eine sehr große Anzahl an Grundgerüsten produzieren, die sich bis auf einige Dekorationen und Verbindungsatome stark ähneln. Diese Informationen sind hochgradig redundant und erschweren eine manuelle Auswertung und Interpretation der Ergebnisse. Die automatisierte Gruppierung von Molekülen auf Basis ihrer Grundgerüste ist bisher noch nicht befriedigend geklärt, was unter anderem auf die unsichere Definition des Grundgerüstbegriffs zurückzuführen ist. (vgl. Abschnitt 2.4) Die in der Literatur veröffentlichten Methoden legen oft einseitig Schwerpunkte auf die funktionellen Gruppen,

wie in verschiedenen Fingerprint-Methoden (z. B. den Daylight Fingerprints[2]), oder auf die Topologie wie im Scaffold Hunter[124].

Mithilfe der zusätzlichen Information, die der Pharmakophor liefert, ließe sich eine automatisierte Gruppierung realisieren, die näher an den in der medizinischen Chemie verwendeten Konzepten liegt. Eine solche Methode würde nur die für den Pharmakophor relevanten funktionellen Gruppen berücksichtigen und sich ansonsten auf die Bewertung beziehungsweise die Klassifizierung der verschiedenen molekularen Topologien konzentrieren. Dies hätte den Vorteil, dass in den Ähnlichkeitsfunktionen nur die Funktionalitäten an den Molekülen, die für die Aktivität relevant sind, beachtet werden müssen. Eine solche Methode könnte bei der manuellen Interpretation der Ergebnisse, die Qsearch liefert, unterstützend eingesetzt werden. Darüber hinaus besitzt dieses Konzept ein großes Potential bei der Gruppierung und Visualisierung von Moleküldatensätzen in Bezug auf eine konkrete Zielstruktur.

Weitergehende intelligente Modifikation von Molekülen Die bisher implementierte Ersetzung von funktionellen Gruppen hat die Ergebnisse von Qsearch bei der Verwendung der BRICS Fragmenträume stark verbessert. Trotzdem bleibt Raum für eine weitergehende Modifikation von Molekülen während der Ausrichtungsphase oder als Nachbearbeitungsschritt. Eine geometrieerhaltende Ersetzung von flexiblen Molekülteilen durch starre Komponenten in einer Nachbearbeitungsphase könnte die Wirkstoffähnlichkeit der Ergebnisse erhöhen und das Hinzufügen von funktionellen Gruppen oder das Verändern von Atomen an Stellen wo ein Pharmakophor nicht erfüllt ist, den Suchraum weiter um relevante Teile erweitern. Die Anwendbarkeit von Teilen dieses Ansatzes wurde bereits in SkelGen[51] gezeigt, wo Heteroatome dynamisch hinzugefügt werden.

In der Diskussion in Abschnitt 6.5.1 wird außerdem ersichtlich, dass der Einsatz hydrophober Pharmakophormerkmale in Qsearch bisher noch nicht uneingeschränkt möglich ist. Insbesondere bei der Verwendung der BRICS-Fragmenträume werden kaum Moleküle erzeugt, die die erwünschten Merkmale aufweisen. Hydrophobe Regionen sind jedoch ein weit verbreitetes Element in vielen publizierten Pharmakophoren (vgl. Abschnitt 6.5.1) – obwohl (oder gerade weil) sie eine sehr unspezifische Beschreibung sind. Daher sollten sie in weiteren Arbeiten Beachtung finden. Denkbar wäre auch hierfür eine weitergehende Entdekoration, die insbesondere nicht-terminale polare funktionelle Gruppen auf geometrisch äquivalente Kohlenstoffgruppen reduziert.

[2] Thor und Merlin; Daylight Chemical Information Systems Inc.: Irvine (CA), USA. `www.daylight.com`

A. Literaturverzeichnis

[1] Klebe, G.: *Wirkstoffdesign.* Spektrum Akademischer Verlag, Heidelberg, 2 Auflage, 2009. 10, 14, 78

[2] Bajorath, J.: *Integration of virtual and high-throughput screening.* Nature Reviews. Drug Discovery, 1(11):882–894, 2002. 10

[3] Stahl, M., W. Guba und M. Kansy: *Integrating molecular design resources within modern drug discovery research: the Roche experience.* Drug Discovery Today, 11(7-8):326–333, 2006. 10, 101

[4] Gellman, S.H.: *Introduction: Molecular Recognition.* Chemical Reviews, 97(5):1231–1232, 1997. 15

[5] Pauling, L., R.B. Corey und H.R. Branson: *The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain.* Proceedings of the National Academy of Sciences of the United States of America, 37(4):205–211, 1951. 15

[6] Arunan, E., G.R. Desiraju, R.A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D.C. Clary, R.H. Crabtree, J.J. Dannenberg, P. Hobza, H.G. Kjaergaard, A.C. Legon, B. Mennucci und D.J. Nesbitt: *Definition of the hydrogen bond (IUPAC Recommendations 2011).* Pure and Applied Chemistry, 83(8):1637–1641, 2011. 16

[7] Boehm, H.J.: *The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure.* Journal of Computer-Aided Molecular Design, 8(3):243–256, 1994. 16, 36

[8] Kumar, S. und R. Nussinov: *Close-range electrostatic interactions in proteins.* Chembiochem : a European Journal of Chemical Biology, 3(7):604–617, 2002. 16

[9] Seebeck, B., I. Reulecke, A. Kamper und M. Rarey: *Modeling of metal interaction geometries for protein-ligand docking.* Proteins, 71(3):1237–1254, 2008. 16

[10] Cerny, J. und P. Hobza: *Non-covalent interactions in biomacromolecules.* Physical Chemistry Chemical Physics : PCCP, 9(39):5291–5303, 2007. 16

[11] Hunter, C.A., J. Singh und J.M. Thornton: *Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins.* Journal of Molecular Biology, 218(4):837–846, 1991. 16

[12] Rella, M., C.A. Rushworth, J.L. Guy, A.J. Turner, T. Langer und R.M. Jackson: *Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors.* Journal of Chemical Information and Modeling, 46(2):708–716, 2006. 17, 95

[13] Kier, L.B.: *Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone.* Molecular Pharmacology, 3(5):487–494, 1967. 17

[14] Kier, L.B.: *MO Theory in Drug Research.* Academic Press, New York, 1971. 17

[15] Van Drie, J.H.: *Monty Kier and the Origin of the Pharmacophore Concept.* Internet Electronic Journal of Molecular Design, 6:271–279, 2007. 17

[16] Wermuth, G., C.R. Ganellin, P. Lindberg und L.A. Mitscher: *Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998).* Pure And Applied Chemistry, 70(5):1129–1143, 1998. 17, 18, 25

[17] Pauli, W.: *Uber den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren.* Zeitschrift fuer Physik, 31(1):765–783, 1925. 18

[18] Spitzer, G.M., M. Heiss, M. Mangold, P. Markt, J. Kirchmair, G. Wolber und K.R. Liedl: *One concept, three implementations of 3D pharmacophore-based virtual screening: distinct coverage of chemical search space.* Journal of Chemical Information and Modeling, 50(7):1241–1247, 2010. 18, 38, 96, 97

[19] Langdon, S.R., P. Ertl und N. Brown: *Bioisosteric Replacement and Scaffold Hopping in Lead Generation and Optimization.* Molecular Informatics, 29(5):366–385, 2010. 19, 20, 65

[20] Reich, H.J. und D.J. Cram: *Multiple Electrophilic Substitution Reactions of [2.2]Paracyclophanes and Interconversions of Polysubstituted Derivatives.* Journal of the American Chemical Society, 91(13):3527, 1969. 19

[21] Bemis, G.W. und M.A. Murcko: *The properties of known drugs. 1. Molecular frameworks.* Journal of Medicinal Chemistry, 39(15):2887–2893, 1996. 19, 21

[22] Boehm, H.J., A. Flohr und M. Stahl: *Scaffold hopping.* Drug Discovery Today: Technologies, 1(3):217–224, 2004. 22, 23

[23] Schneider, G., P. Schneider und S. Renner: *Scaffold-Hopping: How Far Can You Jump?* QSAR & Combinatorial Science, 25(12):1162–1171, 2006. 22, 26

[24] Mauser, H. und W. Guba: *Recent developments in de novo design and scaffold hopping.* Current Opinion in Drug Discovery & Development, 11(3):365–374, 2008. 22, 26, 27

[25] Lessel, U., B. Wellenzohn, M. Lilienthal und H. Claussen: *Searching Fragment Spaces with Feature Trees.* Journal of Chemical Information and Modeling, 49(2):270–279, 2009. 23

[26] Mauser, H. und M. Stahl: *Chemical fragment spaces for de novo design.* Journal of Chemical Information and Modeling, 47(2):318–324, 2007. 23

[27] Lewell, X.Q., D.B. Judd, S.P. Watson und M.M. Hann: *RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry.* Journal of Chemical Information and Computer Sciences, 38(3):511–522, 1998. 23, 27, 102

[28] Schneider, G., O. Clement-Chomienne, L. Hilfiger, P. Schneider, S. Kirsch, H.J. Bohm und W. Neidhart: *Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design.* Angewandte Chemie (International ed. in English), 39(22):4130–4133, 2000. 23, 30

[29] Rarey, M. und M. Stahl: *Similarity searching in large combinatorial chemistry spaces.* Journal of Computer-Aided Molecular Design, 15(6):497–520, 2001. 23, 24, 34, 45

[30] Todorov, N.P. und P.M. Dean: *Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design.* Journal of Computer-Aided Molecular Design, 11(2):175–192, 1997. 23, 28, 29, 33, 49, 98

[31] Degen, J. und M. Rarey: *FlexNovo: structure-based searching in large fragment spaces.* ChemMedChem, 1(8):854–868, 2006. 23, 27, 31, 34, 45

[32] Degen, J., C. Wegscheid-Gerlach, A. Zaliani und M. Rarey: *On the art of compiling and using 'drug-like' chemical fragment spaces.* ChemMedChem, 3(10):1503–1507, 2008. 23, 24, 26, 99

[33] Hartenfeller, M. und G. Schneider: *De novo drug design.* Methods in Molecular Biology (Totowa, NJ, United States), 672:299–323, 2011. 24

[34] Kutchukian, P.S. und E.I. Shakhnovich: *De novo design: balancing novelty and confined chemical space.* Expert Opinion on Drug Discovery, 5(8):789–812, 2010. 24, 26, 27

[35] Boehm, H.J.: *The computer program LUDI: a new method for the de novo design of enzyme inhibitors.* Journal of Computer-Aided Molecular Design, 6(1):61–78, 1992. 25, 27

[36] Gillet, V., A.P. Johnson, P. Mata, S. Sike und P. Williams: *SPROUT: a program for structure generation.* Journal of Computer-Aided Molecular Design, 7(2):127–153, 1993. 25

[37] Clark, D.E., D. Frenkel, S.A. Levy, J. Li, C.W. Murray, B. Robson, B. Waszkowycz und D.R. Westhead: *PRO-LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules.* Journal of Computer-Aided Molecular Design, 9(1):13–32, 1995. 25

[38] Teodoro, M. und I. Muegge: *BIBuilder: Exhaustive Searching for De Novo Ligands.* Molecular Informatics, 30(1):63–75, 2011. 25, 31, 45, 98

[39] Pierce, A.C., G. Rao und G.W. Bemis: *BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease.* Journal of Medicinal Chemistry, 47(11):2768–2775, 2004. 25, 99

[40] Maass, P., T. Schulz-Gasch, M. Stahl und M. Rarey: *Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations.* Journal of Chemical Information and Modeling, 47(2):390–399, 2007. 26, 34, 70

[41] Blum, L.C. und J.-L. Reymond: *970 million druglike small molecules for virtual screening in the chemical universe database GDB-13.* Journal of the American Chemical Society, 131(25):8732–8733, 2009. 26

[42] Bohacek, R.S., C. McMartin und W.C. Guida: *The art and practice of structure-based drug design: a molecular modeling perspective.* Medicinal Research Reviews, 16(1):3–50, 1996. 26

[43] Schneider, P. und G. Schneider: *Collection of Bioactive Reference Compounds for Focused Library Design.* QSAR & Combinatorial Science, 22(7):713–718, 2003. 26, 30, 99

[44] Hartenfeller, M., E. Proschak, A. Schueller und G. Schneider: *Concept of combinatorial de novo design of drug-like molecules by particle swarm optimization.* Chemical Biology & Drug Design, 72(1):16–26, 2008. 26

[45] Fischer, J.R., U. Lessel und M. Rarey: *LoFT: Similarity-Driven Multiobjective Focused Library Design.* Journal of Chemical Information and Modeling, 50(1):1–21, 2010. 26, 98

[46] Schneider, G. und U. Fechner: *Computer-based de novo design of drug-like molecules.* Nature Reviews Drug Discovery, 4(8):649–663, 2005. 27, 31

[47] Zaliani, A., K. Boda, T. Seidel, A. Herwig, C.H. Schwab, J. Gasteiger, H. Claussen, C. Lemmen, J. Degen, J. Parn und M. Rarey: *Second-generation de novo design: a view from a medicinal chemist perspective.* Journal of Computer-Aided Molecular Design, 2009. 27

[48] Schneider, G., M. Hartenfeller, M. Reutlinger, Y. Tanrikulu, E. Proschak und P. Schneider: *Voyages to the (un)known: adaptive design of bioactive compounds.* Trends in Biotechnology, 27(1):18–26, 2009. 27, 45

[49] Rarey, M., B. Kramer, T. Lengauer und G. Klebe: *A fast flexible docking method using an incremental construction algorithm.* Journal of Molecular Biology, 261(3):470–489, 1996. 27

[50] Rarey, M., S. Wefing und T. Lengauer: *Placement of medium-sized molecular fragments into active sites of proteins.* Journal of Computer-Aided Molecular Design, 10(1):41–54, 1996. 28

[51] Todorov, N.P. und P.M. Dean: *A branch-and-bound method for optimal atom-type assignment in de novo ligand design.* Journal of Computer-Aided Molecular Design, 12(4):335–349, 1998. 28, 33, 104

[52] Lloyd, D.G., C.L. Buenemann, N.P. Todorov, D.T. Manallack und P.M. Dean: *Scaffold hopping in de novo design. Ligand generation in the absence of receptor information.* Journal of Medicinal Chemistry, 47(3):493–496, 2004. 29

[53] Todorov, N.P., C.L. Buenemann und I.L. Alberts: *De novo ligand design to an ensemble of protein structures.* Proteins, 64(1):43–59, 2006. 29

[54] Stahl, M., N.P. Todorov, T. James, H. Mauser, H.-J. Boehm und P.M. Dean: *A validation study on the practical use of automated de novo design.* Journal of Computer-Aided Molecular Design, 16(7):459–478, 2002. 29, 78, 79, 95, 101

[55] Dean, P.M., S. Firth-Clark, W. Harris, S.B. Kirton und N.P. Todorov: *SkelGen: a general tool for structure-based de novo ligand design.* Expert Opinion on Drug Discovery, 1(2):179–189, 2006. 29

[56] Guner, O.F.: *Pharmacophore Perception, Development, and Use in Drug Design.* International University Line, La Jolla, CA, 2000. 29

[57] Hert, J., P. Willett, D.J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby und A. Schuffenhauer: *Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures.* Organic & Biomolecular Chemistry, 2(22):3256–3266, 2004. 29

[58] Hert, J., P. Willett, D.J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby und A. Schuffenhauer: *Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures.* Journal of Chemical Information and Computer Sciences, 44(3):1177–1185, 2004. 29

[59] Kirchmair, J., S. Distinto, P. Markt, D. Schuster, G.M. Spitzer, K.R. Liedl und G. Wolber: *How to optimize shape-based virtual screening: choosing the right query and including chemical information.* Journal of Chemical Information and Modeling, 49(3):678–692, 2009. 29

[60] Fechner, U. und G. Schneider: *Flux (1): a virtual synthesis scheme for fragment-based de novo design.* Journal of Chemical Information and Modeling, 46(2):699–707, 2006. 30, 45

[61] Viswanadhan, V.N., A.K. Ghose, G.R. Revankar und R.K. Robins: *Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and*

dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. Journal of Chemical Information and Modeling, 29(3):163–172, 1989. 30

[62] Roche, O., P. Schneider, J. Zuegge, W. Guba, M. Kansy, A. Alanine, K. Bleicher, F. Danel, E.-M. Gutknecht, M. Rogers-Evans, W. Neidhart, H. Stalder, M. Dillon, E. Sjogren, N. Fotouhi, P. Gillespie, R. Goodnow, W. Harris, P. Jones, M. Taniguchi, S. Tsujii, W. von der Saal, G. Zimmermann und G. Schneider: *Development of a virtual screening method for identification of "frequent hitters" in compound libraries.* Journal of Medicinal Chemistry, 45(1):137–142, 2002. 30

[63] Fechner, U. und G. Schneider: *Flux (2): comparison of molecular mutation and crossover operators for ligand-based de novo design.* Journal of Chemical Information and Modeling, 47(2):656–667, 2007. 30

[64] Schuller, A., M. Suhartono, U. Fechner, Y. Tanrikulu, S. Breitung, U. Scheffer, MW Gobel und G. Schneider: *The concept of template-based de novo design from drug-derived molecular fragments and its application to TAR RNA.* Journal of Computer-Aided Molecular Design, 22(2):59–68, 2008. 30, 99

[65] Damewood, J.R.J., C.L. Lerman und B.B. Masek: *NovoFLAP: A ligand-based de novo design approach for the generation of medicinally relevant ideas.* Journal of Chemical Information and Modeling, 50(7):1296–1303, 2010. 30, 49, 98, 101

[66] Grant, J.A., M.A. Gallardo und B.T. Pickup: *A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape.* Journal of Computational Chemistry, 17(14):1653–1666, 1996. 31

[67] Rhodes, N., P. Willett, A. Calvet, J.B. Dunbar und C. Humblet: *CLIP: similarity searching of 3D databases using clique detection.* Journal of Chemical Information and Computer Sciences, 43(2):443–448, 2003. 31

[68] Yadav, V.K., J.-H. Ryu, N. Suda, K.F. Tanaka, J.A. Gingrich, G. Schutz, F.H. Glorieux, C.Y. Chiang, J.D. Zajac, K.L. Insogna, J.J. Mann, R. Hen, P. Ducy und G. Karsenty: *Lrp5 controls bone formation by inhibiting serotonin synthesis in the duodenum.* Cell, 135(5):825–837, 2008. 31

[69] Huang, Y.-yu, M.A. Oquendo, J.M.H. Friedman, L.L. Greenhill, B. Brodsky, K.M. Malone, V. Khait und J.J. Mann: *Substance abuse disorder and major depression are associated with the human 5-HT1B receptor gene*

(HTR1B) G861C polymorphism. Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 28(1):163–169, 2003. 31

[70] Nugiel, D.A., J.R. Krumrine, D.C. Hill, J.R. Damewood Jr., P.R. Bernstein, C.D. Sobotka-Briner, J.W. Liu, A. Zacco und M.E. Pierson: *De Novo Design of a Picomolar Nonbasic 5-HT_{1B}Receptor Antagonist.* Journal of Medicinal Chemistry, 53(4):1876–1880, 2010. 31

[71] Urbaczek, S., A. Kolodzik, J.R. Fischer, T. Lippert, S. Heuser, I. Groth, T. Schulz-Gasch und M. Rarey: *NAOMI: On the Almost Trivial Task of Reading Molecules from Different File formats.* Journal of Chemical Information and Modeling, in Druck, 2011. 34, 37, 41, 50, 57

[72] Clark, M., R.D. Cramer und N. Van Opdenbosch: *Validation of the general purpose tripos 5.2 force field.* Journal of Computational Chemistry, 10(8):982–1012, 1989. 35

[73] Greene, J., S. Kahn, Savoj H., P. Sprague und S. Teig: *Chemical Function Queries for 3D Database Search.* Journal of Chemical Information and Computer Sciences, 34(6):1297, 1994. 38, 95

[74] Leach, A.R., V.J. Gillet, R.A. Lewis und R. Taylor: *Three-dimensional pharmacophore methods in drug discovery.* Journal of Medicinal Chemistry, 53(2):539–558, 2010. 38, 49

[75] Bondi, A.: *van der Waals Volumes and Radii.* The Journal of Physical Chemistry, 68(3):441–451, 1964. 41

[76] De Waterbeemd, H. van, G. Camenisch, G. Folkers und O.A. Raevsky: *Estimation of Caco-2 Cell Permeability using Calculated Molecular Descriptors.* Quantitative Structure-Activity Relationships, 15(6):480–490, 1996. 41

[77] Grant, J.A. und B.T. Pickup: *A Gaussian Description of Molecular Shape.* The Journal of Physical Chemistry, 99(11):3503–3510, 1995. 41, 42, 44

[78] Zhao, Y.H., M.H. Abraham und A.M. Zissimos: *Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds.* The Journal of Organic Chemistry, 68(19):7368–7373, 2003. 41, 42

[79] Irwin, J.J. und B.K. Shoichet: *ZINC–a free database of commercially available compounds for virtual screening.* Journal of Chemical Information and Modeling, 45(1):177–182, 2005. 41

[80] Gasteiger, J., C. Rudolph und J. Sadowski: *Automatic generation of 3D-atomic coordinates for organic molecules.* Tetrahedron Computer Methodology, 3(6, Part 3):537–547, 1990. 42, 44

[81] Schneider, G., M.L. Lee, M. Stahl und P. Schneider: *De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks.* Journal of Computer-Aided Molecular Design, 14(5):487–494, 2000. 45

[82] Dey, F. und A. Caflisch: *Fragment-based de novo ligand design by multiobjective evolutionary optimization.* Journal of Chemical Information and Modeling, 48(3):679–690, 2008. 45

[83] Kirkpatrick, S., C.D. Gelatt und M.P. Vecchi: *Optimization by Simulated Annealing.* Science (New York, N.Y.), 220(4598):671–680, 1983. 45, 50

[84] Oprea, T.I.: *Property distribution of drug-related chemical databases.* Journal of Computer-Aided Molecular Design, 14(3):251–264, 2000. 48

[85] Wolber, G. und T. Langer: *LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters.* Journal of Chemical Information and Modeling, 45(1):160–169, 2005. 49, 85, 86, 95

[86] Hurst, T.: *Flexible 3D Searching: The Directed Tweak Technique.* Journal of Chemical Information and Computer Sciences, 34(1):190–196, 1994. 49

[87] Perola, E. und P.S. Charifson: *Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding.* Journal of Medicinal Chemistry, 47(10):2499–2510, 2004. 49

[88] Weininger, D., A. Weininger und J.L. Weininger: *SMILES. 2. Algorithm for generation of unique SMILES notation.* Journal of Chemical Information and Computer Sciences, 29(2):97–101, 1989. 50

[89] Chan, S.L. und P. Labute: *Training a scoring function for the alignment of small molecules.* Journal of Chemical Information and Modeling, 50(9):1724–1735, 2010. 51, 52, 125

[90] Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov und P.E. Bourne: *The Protein Data Bank.* Nucleic Acids Research, 28(1):235–242, 2000. 51, 85

[91] Leach, A.R.: *Molecular Modelling.* Pearson/Prentice Hall, Harlow, UK, 2 Auflage, 2001. 59, 70, 95

[92] Knuth, D.: *The Art of Computer Programming Volume 2: Seminumerical Algorithms.* Addison-Wesley, Boston (USA), 1997. 62

[93] Spiegler, I. und R. Maayan: *Storage and retrieval considerations of binary data bases.* Information Processing & Management, 21(3):233–254, 1985. 62

[94] Drepper, U.: *What Every Programmer Should Know About Memory.* Linux Weekly News, 2007. 65

[95] Knuth, D.: *The Art of Computer Programming Volume 4A: Combinatorial Algorithms.* Addison-Wesley, Boston (USA), 2011. 65, 67

[96] Brameld, K.A.: *Shape Diversity in Fragment Library Design: Theory and Practice.* In: *FBLD 2010*, Philadelphia, PA, USA, 2010. 70, 71

[97] Lipinski, C.A., F. Lombardo, B.W. Dominy und P.J. Feeney: *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Advanced Drug Delivery Reviews, 46(1-3):3–26, 2001. 73, 95

[98] Vane, J.R.: *Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs.* Nature: New Biology, 231(25):232–235, 1971. 78

[99] Dannhardt, G. und S. Laufer: *Structural approaches to explain the selectivity of COX-2 inhibitors: is there a common pharmacophore?* Current Medicinal Chemistry, 7(11):1101–1112, 2000. 78

[100] Kurumbail, R.G., A.M. Stevens, J.K. Gierse, J.J. McDonald, R.A. Stegeman, J.Y. Pak, D. Gildehaus, J.M. Miyashiro, T.D. Penning, K. Seibert, P.C. Isakson und W.C. Stallings: *Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents.* Nature, 384(6610):644–648, 1996. 79

[101] Stahl, M. und M. Rarey: *Detailed analysis of scoring functions for virtual screening.* Journal of Medicinal Chemistry, 44(7):1035–1042, 2001. 79

[102] Garg, R., A. Kurup, S.B. Mekapati und C. Hansch: *Cyclooxygenase (COX) inhibitors: a comparative QSAR study.* Chemical Reviews, 103(3):703–732, 2003. 82

[103] Sloma, I., X. Jiang, A.C. Eaves und C.J. Eaves: *Insights into the stem cells of chronic myeloid leukemia.* Leukemia : Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 24(11):1823–1833, 2010. 84

[104] Levitzki, A. und E. Mishani: *Tyrphostins and other tyrosine kinase inhibitors.* Annual Review of Biochemistry, 75:93–109, 2006. 84

[105] Nagar, B., O. Hantschel, M.A. Young, K. Scheffzek, D. Veach, W. Bornmann, B. Clarkson, G. Superti-Furga und J. Kuriyan: *Structural basis for the autoinhibition of c-Abl tyrosine kinase.* Cell, 112(6):859–871, 2003. 85

[106] Nogrady, T. und D.F. Weaver: *Medicinal Chemistry.* Oxford University Press, New York, NY, USA, 3 Auflage, 2005. 88

[107] Lindberg, J., S. Sigurdsson, S. Lowgren, H.O. Andersson, C. Sahlberg, R. Noreen, K. Fridborg, H. Zhang und T. Unge: *Structural basis for the inhibitory efficacy of efavirenz (DMP-266), MSC194 and PNU142721 towards the HIV-1 RT K103N mutant.* European Journal of Biochemistry / FEBS, 269(6):1670–1677, 2002. 92

[108] Perno, C.-F., G. Moyle, C. Tsoukas, W. Ratanasuwan, J. Gatell und M. Schechter: *Overcoming resistance to existing therapies in HIV-infected patients: the role of new antiretroviral drugs.* Journal of Medical Virology, 80(4):565–576, 2008. 92

[109] Waterbeemd, H. van de, R.E. Carter, G. Grassy, H. Kubinyi, Y.C. Martin, M.S. Tute und P. Willet: *Glossary of terms used in computational drug design (IUPAC Recommendations 1997).* Pure And Applied Chemistry, 69(5):1137–1152, 1997. 94

[110] Bleicher, K.H., H.-J. Bohm, K. Muller und A.I. Alanine: *Hit and lead generation: beyond high-throughput screening.* Nature Reviews. Drug Discovery, 2(5):369–378, 2003. 95

[111] Roche, O. und R.M. Rodriguez Sarmiento: *A new class of histamine H3 receptor antagonists derived from ligand based design.* Bioorganic & Medicinal Chemistry Letters, 17(13):3670–3675, 2007. 95

[112] Wolber, G., T. Seidel, F. Bendix und T. Langer: *Molecule-pharmacophore superpositioning and pattern matching in computational drug design.* Drug Discovery Today, 13(1-2):23–29, 2008. 95, 96

[113] Kauzmann, W.: *Some factors in the interpretation of protein denaturation.* Advances in Protein Chemistry, 14:1–63, 1959. 97

[114] Reulecke, I., G. Lange, J. Albrecht, R. Klein und M. Rarey: *Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function.* ChemMedChem, 3(6):885–897, 2008. 97

[115] Friesner, R.A., R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin und D.T. Mainz: *Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.* Journal of Medicinal Chemistry, 49(21):6177–6196, 2006. 97

[116] Allen, F.H.: *The Cambridge Structural Database: a quarter of a million crystal structures and rising.* Acta Crystallographica. Section B, Structural Science, 58(3):380–388, 2002. 97

[117] Klebe, G. und T. Mietzner: *A fast and efficient method to generate biologically relevant conformations.* Journal of Computer-Aided Molecular Design, 8(5):583–606, 1994. 97

[118] Brameld, K.A., B. Kuhn, D.C. Reuter und M. Stahl: *Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis.* Journal of Chemical Information and Modeling, 48(1):1–24, 2008. 97

[119] Griewel, A., O. Kayser, J. Schlosser und M. Rarey: *Conformational Sampling for Large-Scale Virtual Screening: Accuracy versus Ensemble Size.* Journal of Chemical Information and Modeling, 49:2303–2311, 2009. 97

[120] Morley, S.D., R.J. Abraham, I.S. Haworth, D.E. Jackson, M.R. Saunders und J.G. Vinter: *COSMIC(90): an improved molecular mechanics treatment*

of hydrocarbons and conjugated systems. Journal of Computer-Aided Molecular Design, 5(5):475–504, 1991. 98

[121] Pearce, B.C., D.R. Langley, J. Kang, H. Huang und A. Kulkarni: *E-novo: an automated workflow for efficient structure-based lead optimization.* Journal of Chemical Information and Modeling, 49(7):1797–1809, 2009. 99

[122] Pegg, S.C., J.J. Haresco und I.D. Kuntz: *A genetic algorithm for structure-based de novo design.* Journal of Computer-Aided Molecular Design, 15(10):911–933, 2001. 99

[123] Makhija, M.T., R.T. Kasliwal, V.M. Kulkarni und N. Neamati: *De novo design and synthesis of HIV-1 integrase inhibitors.* Bioorganic & Medicinal Chemistry, 12(9):2317–2333, 2004. 99

[124] Wetzel, S., K. Klein, S. Renner, D. Rauh, T.I. Oprea, P. Mutzel und H. Waldmann: *Interactive exploration of chemical space with Scaffold Hunter.* Nature Chemical Biology, 5(8):581–583, 2009. 104

B. Benutzerhandbuch

Qsearch kann entweder über eine grafischen Benutzerschnittstelle oder über die Kommandozeile ausgeführt werden. Die Konfigurationsmöglichkeiten für beide Modi werden im folgenden dargelegt.

B.1. Grafische Benutzerschnittstelle

Die grafische Benutzerschnittstelle ist der Standardmodus von Qsearch. Sie wird geöffnet, wenn das Programm ohne Parameter aufgerufen wird. (Abbildung B.1.1) Die einzelnen Interaktionsmöglichkeiten sind wie folgt:

B.1.1. Menüpunkt „Actions“

Unter dem Menüpunkt „Actions“ gibt es vier Unterpunkte, die folgende Bedeutungen haben:

Search Molecules Ein Dialog wird geöffnet, um eine Suche aufzusetzen. Die Angabe eines Fragmentraums und eines Pharmakophors ist obligatorisch. Zusätzlich lassen sich noch die Suche, die Ausrichtung und die schnelle Ablehnung konfigurieren. Nach der Bestätigung mit „OK“ öffnet sich die Suchkontrolle.

Grow Molecules Ein Dialog, der mit dem von „Search Molecules“ vergleichbar ist, wird geöffnet. Er erlaubt das Aufsetzen eines Fragmenterweiterungsexperiments. Zusätzlich zu den bei „Search Molecules“ angegebenen Informationen muss noch ein Startfragment ausgewählt werden.

Write Commands Die aktuelle Konfiguration wird in eine Datei für den Start des Kommandozeilenmodus geschrieben. Hierfür muss vorher entweder über „Search Molecules“ oder „Grow Molecules“ eine Konfiguration erstellt worden sein.

Quit Beendet das Programm.

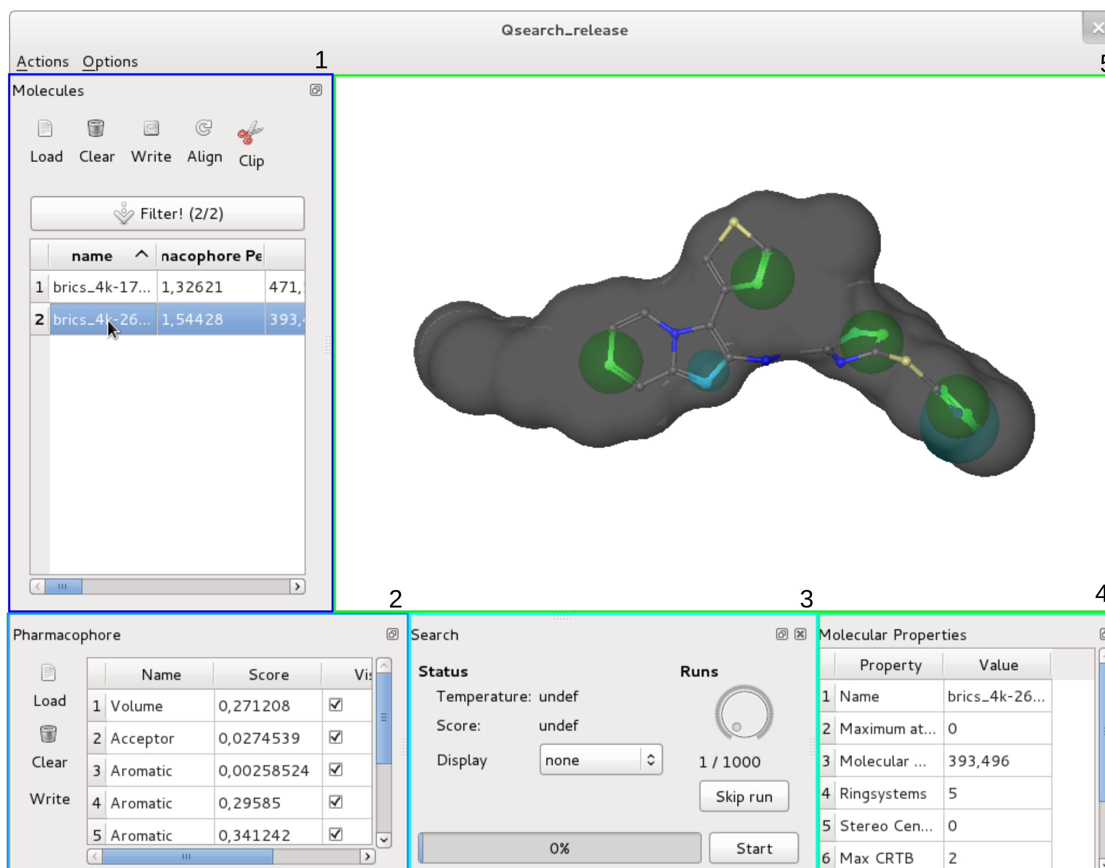


Abbildung B.1.1.: Grafische Benutzerschnittstelle nach dem Start. 1: Molekül-Liste 2: Pharmakophoreigenschaften 3: Suchkontrolle 4: Moleküleigenschaften 5: 3D-Anzeige

B.1.2. Menüpunkt „Options“

Hier lassen sich Optionen für die 3D-Darstellung verändern. Es ist möglich, Wasserstoffatome auszublenden, und die Hintergrundfarbe zu verändern. Darüber hinaus sind hier Versionsinformationen des Programms einsehbar.

B.1.3. Molekül-Liste

In der Molekül-Liste wird eine Menge von Molekülen verwaltet. Wenn eine Suche läuft, werden die Ergebnismoleküle hier angezeigt. Mit der Liste kann über verschiedene Methoden interagiert werden:

Klicken eines Moleküls Wenn ein Molekül angeklickt wird, wird es im 3D Fenster angezeigt, und Informationen zu dem Molekül werden in den Moleküleigenschaften berechnet und dargestellt. Zusätzlich werden Strafterme für die einzelnen Pharmakophormerkmale berechnet und in den Pharmakophoreigenschaften dargestellt.

Load Über diesen Knopf können Moleküle aus einer externen Datei in die Liste geladen werden.

Clear Über diesen Knopf können Moleküle aus der Liste entfernt werden. Nach einem Klick wird man vor die Wahl gestellt, entweder die aktuell ausgewählten Moleküle oder die gesamte Liste zu löschen.

Write Über diesen Knopf können Moleküle in eine Datei geschrieben werden. Nach einem Klick wird man vor die Wahl gestellt, entweder die aktuell ausgewählten Moleküle oder die sämtliche Moleküle zu schreiben.

Align Über diesen Knopf wird das aktuell ausgewählte Molekül an dem Pharmakophor ausgerichtet.

Clip Über diesen Knopf wird der Entdekorierungsschritt auf das aktuell ausgewählte Molekül angewendet.

Filter Über diesen Knopf kann ein Filter im NAOMI Format auf die Molekül-Liste angewendet werden. Auf dem Knopf sind zwei Zahlen in einer Klammer angezeigt. Die erste Zahl zeigt an, wie viele Moleküle dem Filter entsprechen und die zweite Zahl zeigt an, wie viele Moleküle insgesamt in der Liste vorhanden sind.

B.1.4. Pharmakophoreigenschaften

In den Pharmakophoreigenschaften werden Informationen zum aktuell geladenen Pharmakophor angezeigt. In der Tabelle auf der rechten Seite sind die einzelnen Merkmale der Pharmakophors aufgelistet. Wenn ein Molekül aktiviert ist, werden die Strafwerte der einzelnen Merkmale angezeigt. Zusätzlich können die Merkmale in der 3D-Anzeige ein- und ausgeblendet werden.

Mit den Eigenschaften kann wie folgt interagiert werden:

Load Über diesen Knopf kann ein Pharmakophor geladen werden.

Clear Über diesen Knopf kann der geladene Pharmakophor gelöscht werden.

Write Über diesen Knopf kann der aktuelle Pharmakophor in eine Datei geschrieben werden.

B.1.5. Suchkontrolle

Dieses Fenster wird erst sichtbar, nachdem eine Suche aufgesetzt worden ist. Es zeigt Informationen zum Verlauf des Algorithmus an, und erlaubt eine Interaktion mit dem Suchalgorithmus:

Display Über diesen Knopf lässt sich die Suche in der 3D-Anzeige verfolgen. Nach einem Klick kann man zwischen folgenden Optionen auswählen: „none“ – keine Moleküle anzeigen, „last tested“ – nach jedem Schritt das Molekül, das am Pharmakophor ausgerichtet wurde, anzeigen, „last accepted“ – Moleküle nur anzeigen, wenn sie im Algorithmus akzeptiert wurden.

Skip run Über diesen Knopf lässt sich der aktuelle Lauf überspringen. Der Algorithmus fängt dann mit dem nächsten Lauf an.

Start/Stopp Über diesen Knopf kann die Suche gestartet und pausiert werden.

B.1.6. 3D-Anzeige

In der 3D-Anzeige können Pharmakophore und Moleküle angezeigt werden. Die 3D-Anzeige lässt folgende Aktionen zu:

Zoomen Entweder über das Mausrad oder Shift+rechte Maustaste gedrückt halten, und die Maus nach links oder rechts bewegen.

Rotieren (X,Y-Achse) Rechte Maustaste gedrückt halten und die Maus bewegen.

Rotation (Z-Achse) Shift+linke Maustaste drücken und die Maus bewegen.

Translation Mittlere Maustaste drücken und die Maus bewegen.

B.2. Kommandozeilenmodus

Qsearch lässt sich für die Ausführung auf einem Rechencluster auf der Kommandozeile ausführen. Hierfür muss zuerst in der grafischen Benutzerschnittstelle unter „Actions/Write Commands“ eine Konfigurationsdatei erstellt worden sein. Die Kommandozeilenparameter sind wie folgt:

--commands <DATEI> (obligatorisch) Angabe zu verwendenden Konfigurationsdatei für den Lauf.

--output <DATEI> (optional) Angabe der Datei, in die die Ergebnisse geschrieben werden sollen.

Durch einen Aufruf mit einer Konfigurationsdatei startet das Programm und beendet sich nach der angegebenen Anzahl an Durchläufen. Es kann vorzeitig mit der Tastenkombination STRG+C beendet werden.

C. Volumenbeiträge der NAOMI Atomtypen

Name	Volumenbeitrag [Å³]	Name	Volumenbeitrag [Å³]
O(100-)	11,542271	S(010)	20,069154
C(020)	10,803928	H(100)	5,125386
S(200)-Aro	19,246438	F(100)	10,269649
S(110+)-Aro	16,393415	N(110)-Aro	8,985621
I(300)	27,629973	S(020)	19,260432
O(110+)	8,426245	S(110+)	15,883807
O(200)-Aro	9,117693	P(300)	16,137557
O(200)-Planar	8,279259	O(100-)-Conjugated	11,341938
Cl(100)	19,324749	N(001)	10,584680
N(300)-Aro	6,869936	N(300)	6,944602
N(101+)	8,512860	Br(100)	23,743314
N(200-)	9,778527	S(200)-Planar	19,010300
C(300-)	10,027208	O(200)	8,528727
P(500)	11,223993	S(220)	12,151092
N(210+)	6,536502	O(-)-Deloc	10,675000
P(400+)	9,795130	S(300+)	15,399694
S(100-)	21,424138	N(010-)	12,008186
N(400+)	4,034228	N(210+)-Aro	5,941945
C(300+)	6,691167	C(101)	11,113931
N(+)-Aro-Deloc	7,033592	P(310)	13,578819
I(100)	29,929109	C(400)	6,504041
S(200)	19,284731	N(020+)	6,807981
C(210)-Aro	8,569893	C(210)	8,597978
N(300)-Planar	6,644852	N(+)-Deloc	6,293964
O(110+)-Aro	6,560777	N(110)	8,867856
S(210)	15,462231	O(010)	10,364912

Tabelle C.1.: NAOMI-Atomtypen und ihre berechneten Volumenbeiträge.

D. Anhang für die Parametrisierung der Ausrichtung

117423A_a_10	50878A_a_1	53392B_6	56514A_a_3
48119A_a_5	50893A_a_30	53881A_a_a_2	56695B_6
48266A_a_2	50943B_1	55629A_b_1	63621A_1
48276A_8	51388A_a_1	55833A_a_67	69423A_2
48519A_a_1	51397A_a_2	55876A_c_2	69548A_2
48521A_c_1	51436A_b_3	55876A_d_b_1	81362A_b_5
48554A_b_b_8	51528A_b_1	55924A_c_2	88855A_a_1
49486A_1	51715A_a_1	56129A_a_7	88855A_c_c_1
50550A_a_1	52735A_1	56129A_b_1	89151A_1
50671A_b_2	53189A_3	56245A_1	90036F_8

Tabelle D.1.: Komplexe aus dem Datensatz von Chan und Labute[89], die zur Parametrisierung der Ausrichtung verwendet wurden.

T_{start}	α	n	„kalt“	max. Rot.	max. Trans.	max. Tors.	% Rot.	% Trans.	% Tors.
20,0	0,986	700	25%	15°	0,4Å	15°	37,5%	12,5%	50,0%
20,0	0,990	700	50%	15°	0,4Å	15°	25,0%	8,3%	66,7%
20,0	0,994	700	75%	15°	0,4Å	15°	50,0%	16,7%	33,3%
20,0	0,990	1000	25%	10°	0,4Å	15°	37,5%	12,5%	50,0%
20,0	0,993	1000	50%	10°	0,4Å	15°	25,0%	8,3%	66,7%
20,0	0,995	1000	75%	10°	0,4Å	15°	50,0%	16,7%	33,3%
20,0	0,992	1300	25%	5°	0,4Å	15°	37,5%	12,5%	50,0%
20,0	0,995	1300	50%	5°	0,4Å	15°	25,0%	8,3%	66,7%
20,0	0,997	1300	75%	5°	0,4Å	15°	50,0%	16,7%	33,3%

(a) Simulated Annealing (b) Konformationsänderungen

Tabelle D.2.: Parametersätze für die Parametrisierungsstudie im ersten Lauf.

T_{start}	α	n	„kalt“	max. Rot.	max. Trans.	max. Tors.	% Rot.	% Trans.	% Tors.
10,0	0,993	1300	25%	10°	0,4Å	10°	45,0%	15,0%	40,0%
20,0	0,990	1300	25%	10°	0,4Å	15°	45,0%	15,0%	40,0%
30,0	0,989	1300	25%	10°	0,4Å	20°	45,0%	15,0%	40,0%
10,0	0,991	1300	20%	10°	0,4Å	10°	50,0%	16,7%	33,3%
20,0	0,988	1300	20%	10°	0,4Å	15°	50,0%	16,7%	33,3%
30,0	0,986	1300	20%	10°	0,4Å	20°	50,0%	16,7%	33,3%
10,0	0,988	1300	15%	10°	0,4Å	10°	53,6%	17,9%	28,6%
20,0	0,984	1300	15%	10°	0,4Å	15°	53,6%	17,9%	28,6%
30,0	0,982	1300	15%	10°	0,4Å	20°	53,6%	17,9%	28,6%

(a) Simulated Annealing (b) Konformationsänderungen

Tabelle D.3.: Parametersätze für die Parametrisierungsstudie im zweiten Lauf.

T_{start}	α	n	„kalt“
20,0	0,990	1300	25%
20,0	0,988	1300	20%
20,0	0,984	1300	15%
20,0	0,991	1400	25%
20,0	0,989	1400	20%
20,0	0,985	1400	15%
20,0	0,992	1500	25%
20,0	0,990	1500	20%
20,0	0,986	1500	15%

(a) Simulated Annealing

max. Rot.	max. Trans.	max. Tors.	% Rot.	% Trans.	% Tors.
15°	0,2Å	15°	35,7%	35,7%	28,6%
15°	0,4Å	15°	35,7%	35,7%	28,6%
15°	0,6Å	15°	35,7%	35,7%	28,6%
15°	0,2Å	15°	47,6%	23,8%	28,6%
15°	0,4Å	15°	47,6%	23,8%	28,6%
15°	0,6Å	15°	47,6%	23,8%	28,6%
15°	0,2Å	15°	53,6%	17,6%	28,6%
15°	0,4Å	15°	53,6%	17,6%	28,6%
15°	0,6Å	15°	53,6%	17,6%	28,6%

(b) Konformationsänderungen

Tabelle D.4.: Parametersätze für die Parametrisierungsstudie im dritten Lauf.

T_{start}	α	n	„kalt“
15,0	0,981	1300	25%
20,0	0,984	1300	20%
25,0	0,986	1300	15%
15,0	0,989	1400	25%
20,0	0,988	1400	20%
25,0	0,987	1400	15%
15,0	0,991	1500	25%
20,0	0,990	1500	20%
25,0	0,989	1500	15%

(a) Simulated Annealing

max. Rot.	max. Trans.	max. Tors.	% Rot.	% Trans.	% Tors.
10°	0,4Å	10°	50,0%	16,6%	33,3%
10°	0,4Å	15°	50,0%	16,6%	33,3%
10°	0,4Å	20°	50,0%	16,6%	33,3%
10°	0,4Å	10°	50,0%	16,6%	33,3%
10°	0,4Å	15°	50,0%	16,6%	33,3%
10°	0,4Å	20°	50,0%	16,6%	33,3%
10°	0,4Å	10°	50,0%	16,6%	33,3%
10°	0,4Å	15°	50,0%	16,6%	33,3%
10°	0,4Å	20°	50,0%	16,6%	33,3%

(b) Konformationsänderungen

Tabelle D.5.: Parametersätze für die Parametrisierungsstudie im vierten Lauf.

E. Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegeben Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt. Es wurde an keinem anderen Fachbereich ein Antrag auf Eröffnung eines Promotionsverfahrens gestellt.

Köln, 8. Dezember 2011

(Tobias Lippert)